International Journal of Computer Vision
https://doi.org/10.1007/s11263-021-01520-5

Assignment Flow for Order-Constrained OCT Segmentation

Dmitrij Sitenko[1] · **Bastian Boll**[1] · **Christoph Schnörr**[2]

Received: 7 January 2021 / Accepted: 18 August 2021

Abstract
At the present time optical coherence tomography (OCT) is among the most commonly used non-invasive imaging methods for the acquisition of large volumetric scans of human retinal tissues and vasculature. The substantial increase of accessible highly resolved 3D samples at the optic nerve head and the macula is directly linked to medical advancements in early detection of eye diseases. To resolve decisive information from extracted OCT volumes and to make it applicable for further diagnostic analysis, the exact measurement of retinal layer thicknesses serves as an essential task be done for each patient separately. However, manual examination of OCT scans is a demanding and time consuming task, which is typically made difficult by the presence of tissue-dependent speckle noise. Therefore, the elaboration of automated segmentation models has become an important task in the field of medical image processing. We propose a novel, purely data driven *geometric approach to order-constrained 3D OCT retinal cell layer segmentation* which takes as input data in any metric space and can be implemented using only simple, highly parallelizable operations. As opposed to many established retinal layer segmentation methods, we use only locally extracted features as input and do not employ any global shape prior. The physiological order of retinal cell layers and membranes is achieved through the introduction of a smoothed energy term. This is combined with additional regularization of local smoothness to yield highly accurate 3D segmentations. The approach thereby systematically avoid bias pertaining to global shape and is hence suited for the detection of anatomical changes of retinal tissue structure. To demonstrate its robustness, we compare two different choices of features on a data set of manually annotated 3D OCT volumes of healthy human retina. The quality of computed segmentations is compared to the state of the art in automatic retinal layer segmention as well as to manually annotated ground truth data in terms of mean absolute error and Dice similarity coefficient. Visualizations of segmented volumes are also provided.

Keywords Assignment flow · Assignment manifold · Optical coherence tomography · Information geometry · Covariance descriptor

1 Introduction

1.1 Overview, Motivation

Optical coherence tomography (OCT) is a non-invasive imaging technique which measures the intensity response of back scattered light from millimeter penetration depth. Here we consider its use in ophthalmology as a means of aquiring high-resolution volume scans of human retina in vivo to understand eye functionalities. Figure 1 gives an overview of relevant anatomy. OCT devices record multiple two-dimensional B-scans in rapid succession and combine them to a single volume in a subsequent alignment step. Taking an OCT scan only takes multiple seconds to few minutes and can help detect symptoms of pathological conditions such as glaucoma, diabetes, multiple sclerosis or age-related

Communicated by Zeynep Akata.

✉ Dmitrij Sitenko
dmitrij.sitenko@iwr.uni-heidelberg.de

Bastian Boll
bastian.boll@iwr.uni-heidelberg.de

Christoph Schnörr
schnoerr@math.uni-heidelberg.de

[1] Image and Pattern Analysis Group (IPA) and Heidelberg Collaboratory for Image Processing (HCI), Heidelberg University, Heidelberg, Germany

[2] Image and Pattern Analysis Group, Heidelberg University, Heidelberg, Germany

Published online: 03 September 2021

macular degeneration. The relative ease of data acquisition also enables to use multiple OCT volume scans of a single patient over time to track the progression of a pathology or quantify the success of therapeutic treatment. As a consequence of the technological progress in OCT imaging which was made over the past few decades since its invention by Huang et al. (1991), more expertise for extraction of manual annotations is required which in the presence of big volumetric data sets is difficult to access.

To better leverage the availability of retinal OCT data in both clinical settings and empirical studies, much work is focused on the analysis of appropriate automatic feature extraction techniques. In particular, the access to such methods is especially crucial for achieving enhanced effectiveness of existing quantitative retinal multi cell layer segmentation approaches, and for increasing their clinical potential in real life applications, such as detection of fluid regions and reconstruction of vascular structures. The difficulty of these tasks lies in the challenging signal-to-noise ratio which is influenced by multiple factors including physical eye movement during registration and the presence of speckle noise.

In this paper, we extend the assignment flow approach proposed in Åström et al. (2017) for labeling data on graphs to automatic cell layer segmentation in OCT data. After a feature extraction step, each voxel is labeled by smoothing local layer decisions and jointly leveraging a global geometric invariant—the natural order of cell layers along the vertical axis of each B-scan, as shown in the second row of Fig. 2. We are able to produce high-quality segmentations of OCT volumes by using *local* features as input for a purpose-built assignment flow variant which serves to incorporate global context in a controlled way. This is in contrast to common machine learning approaches which use essentially full B-scans as input.

The empirical success of deep learning methods is driven by the striking ability of deep networks to discover informative features which capture even very subtle patterns in data. However, despite their apparent expressiveness, such features are notoriously hard to interpret by humans. While neural networks often generalize surprisingly well to unseen data, their lack of interpretability makes it hard to anticipate or otherwise reason about specific failure cases. This is particularly relevant in medical applications because deep networks may produce predictions which appear plausible even in cases where they fail to generalize. Additionally, the aquisition of high-quality labeled data for training is laborious and may require the expertise of skilled medical professionals such that data availibility is limited compared to other problem domains. We propose to localize the influence of feature extraction on the segmentation process by limiting field of view. Consequently, the used features are semantically weaker than the ones computed by competing deep learning methods. However, we still achieve state of the art performance by leveraging domain knowledge. In our pipeline, ambiguities in local features are resolved by regularizing to achieve local regularity as well as physiological cell layer ordering.

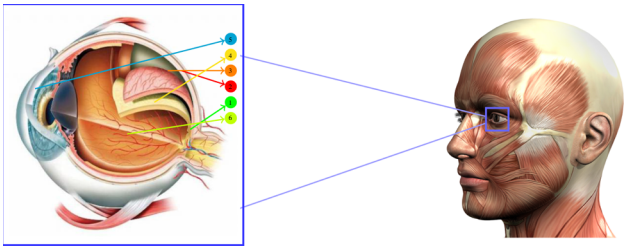


Fig. 1 Schematic illustration of human eye functionality designed by Kjpargeter (n.d): Light enters the Cornea 5 and passes through the vitreous humour 6 towards the retina 4 and choroid 3 which are located around the fovea 1

Our segmentation approach is a smooth image labeling algorithm based on geometric numerical integration on an elementary statistical manifold. It can work with input data from any metric space, making it agnostic to the choice of feature extraction and suitable as plug-in replacement in diverse pipelines. In addition to respecting the natural order of cell layers, the proposed segmentation process has a high amount of built-in parallelism such that modern graphics acceleration hardware can easily be leveraged. We compare the effectiveness of our novel approach between a selection of input features ranging from traditional covariance descriptors to convolutional neural networks. Figure 2 shows a typical volume segmentation computed by the proposed method. It illustrates how local ambiguity is caused by similar signal intensity and visual appearance of some layers and exacerbated by speckle noise. This ambiguity in local features is systematically resolved by leveraging the domain knowledge of local smoothness and global physiological layer order.

1.2 Related Work

Effective segmentation of OCT volumes is a very active area of research. Here, we briefly review the current state of the art approaches originating from the broad research fields of graphical models, variational methods and machine learning.

1.2.1 Graphical Models

The first mathematical access to the problem is provided by the theory of graphical models which transforms the segmentation task into an optimization problem with hard pairwise interaction constraints between voxels. Starting with Kang et al. (2006) and Haeker et al. (2007), simultaneous retina layer detection attempts were made by finding an s-t minimum graph cut. Garvin et al. (2009) further extended this approach with a shape prior modeling layer boundaries. The

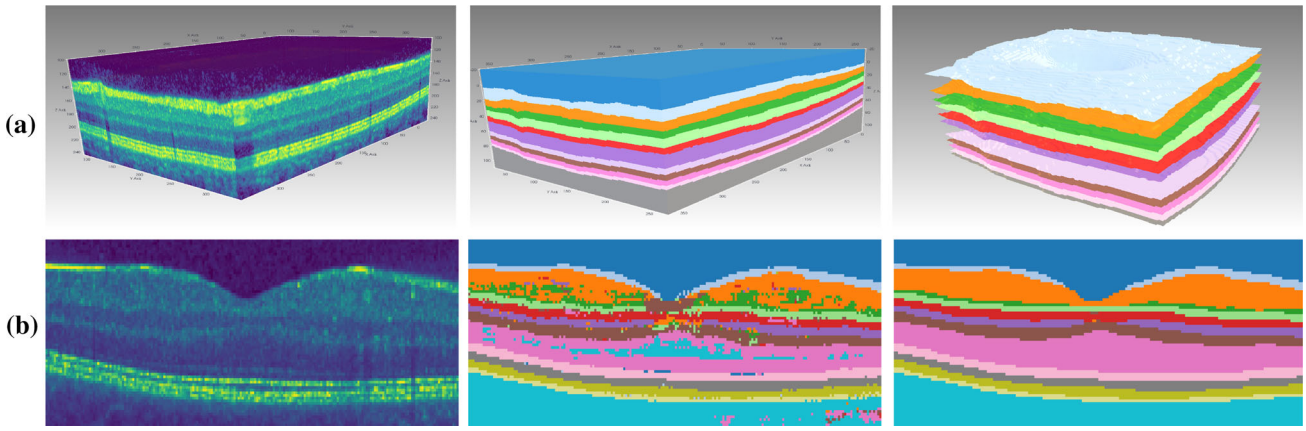


Fig. 2 **a** *Left* Normalized view on a 3D OCT volume scan dimension $512 \times 512 \times 256$ of healthy human retina with ambiguous locations of layer boundaries. *Middle* The resulting segmentation of 11 layers displaying the order preserving labeling of the proposed approach. *Right* Boundary surfaces between different segmented cell layers are illustrated. **b** Typical result of the proposed segmentation approach for a single B-scan of healthy retina. *Left* raw OCT input data. *Middle* segmentation by locally selecting the label with maximum score for each voxel after feature extraction. *Right* segmentation by the proposed assignment flow approach using the same extracted features

methods benefit from low computational complexity, but are lacking of robustness in the presence of speckle and therefore require additional preprocessing steps. Along this line of reasoning, Antony et al. (2010) used a two stage segmentation process by applying anisotropic diffusion in a preprocessing step and consequently segmenting outer retina layers using graphical models. Similarly, Kafieh et al. (2013) proposed to use specific distances based on diffusion maps which are computed by coarse graining the original graph. However, increased performance for noisy OCT data gained by regularizing in this way comes at the cost of introducing bias in the preprocessing step which in turn inpairs robustness in settings with medical pathologies.

Motivated by Song et al. (2013), Dufour et al. (2013) comes up with a circular shape prior for segmentation of 6 retinal layers by incorporating soft constraints which are more suitable for the robust detection of pathological retina structures. Chiu et al. (2015) relies on a graphical model approach as a postprocessing step after applying a supervised kernel regression classification with features extracted according to Quellec et al. (2010). Rathke et al. (2014) reduced the overall complexity by a parallelizable segmentation approach based on probabilistic graphical models with global low-rank shape prior representing interacting retina tissues surfaces. While the *global* shape prior works well for non-pathological OCT data, it cannot be adapted to the broad range of variations caused by *local pathological* structure resulting in a inherent limitation of this approach. Here we refer to Rathke et al. (2017) for possible adaption of the probabilistic approach (Rathke et al. 2014) to pathological retina detection.

1.2.2 Variational Methods

Another category of layer detection methods focus on minimizing an energy functional to express the quantity of interest as the solution to an optimization problem. To this class of methods for retina detection level set approaches have proven to be particularly suitable by encoding each retina layer as the zero level sets of a certain functional. Yazdanpanah et al. (2011) introduces a level set method for minimizing an active contour functional supported by a multiphase model presented in Chan and Vese (2001) as circular shape prior, to avoid limitations of hard constraints as opposed to graphical model proposed by Garvin et al. (2009). Duan et al. (2015) suggests the approach to model layer boundaries with a mixture of Mumford Shah and Vese and Osher functionals by first preprocessing the data in the Fourier domain. A capable level set approach for joint segmentation of pathological retina tissues was reported in the work of Novosel et al. (2017). However, due to the involved hierarchical optimization, their method is computationally expensive. One common downside of the above algorithms are their inherent limitations to only include local notions of layer ordering, making their extension to cases with pathologically caused retina degeneracy a difficult task.

1.2.3 Machine Learning

Much recent work has focused on the use of deep learning to address the task of cell layer segmentation in a purely data driven way. The U-net architecture (Ronneberger et al. 2015) has proven influential in this domain because of its good

predictive performance in settings with limited availability of training data. Multiple modifications of U-net have been proposed to specifically increase its performance in OCT applications (Roy et al. 2017; Liu et al. 2019). The common methods largely rely on convolutional neural networks to predict layer segmentations for individual B-scans which are subsequently combined to full volumes. These methods have also been used as part of a two-stage pipeline where additional prior knowledge such as local regularity and global order of cell layers along a spatial axis is incorporated through graph-based methods (Fang et al. 2017) or a second machine learning component (He et al. 2019).

1.3 Contribution, Organization

We propose a geometric assignment approach to retinal layer segmentation. By leveraging a continuous characterization of layer ordering, our method is able to simultaneously perform local regularization and incorporate the global topological ordering constraint in a *single smooth* labeling process. The segmentation is computed from a distance matrix containing pairwise distances between data for each voxel and prototypical data for each layer in some feature space. This highlights the ability to extract features from raw OCT data in a variety of different ways and to use the proposed segmentation as a plug-in replacement for other graph-based methods.

As a result of the proposed method, it becomes possible to compute high-quality cell layer segmentations of OCT volumes by using only local features for each voxel. This is in contrast to competing deep learning approaches which explicitly aim to incorporate as much global context into the feature extraction process as possible. The exclusive use of local features combats bias introduced through limited data availability in training and makes incorporation of three-dimensional information easily possible without limiting runtime scalability. To demonstrate this, we implement two feature extraction approaches. The first is based on identifying each voxel with a covariance descriptor and finding prototypical descriptors as cluster centers. For each voxel, Riemannian distances to the prototypical descriptors are used as input for subsequent segmentation. The second is based on training a relatively shallow convolutional neural network to classify small voxel patches of raw OCT data. Predicted class scores for each voxel are subsequently used as input for the proposed segmentation method.

The final pipeline thus comprises a preliminary *feature extraction* step (summarized in Sect. 5.2) which yields local data to subsequently be labeled in a regularized fashion by the proposed *ordered assignment flow* (Definition 2).

It enables robust cell layer segmentation for raw OCT volumes at scale, labeling an entire OCT volume in the time frame between 30 s and several minutes on a single GPU and in general leads to increased performance in the case of more informative features. This is without using any prior knowledge other than local regularity and order of cell layers. In particular, no global shape prior is used thus making our proposed approach suited for retina detection in OCT volumes with observable pathological patterns.

Our paper considerably elaborates the conference version (Sitenko et al. 2020) in the following ways. We extended the discussion of related work and added descriptions of two reference methods to make the paper more self-contained. The mechanism we use to promote topological layer ordering through regularization is based on a generalized notion of order preservation restated in Definition 1. In the present work, we motivate this notion by examining a related discrete graphical model in Proposition 1. Furthermore, regarding the choice of covariance descriptors (Tuzel et al. 2006) for retinal tissue representation we extensively discussed the impact of retrieving prototypical descriptors by approximating Riemannian distance via divergence functions. Accordingly, we provided a detailed qualitative performance evaluation in terms of the labeling accuracy and computational efficiency by comparing to the alternative Riemannian mean retrieval approach (Bini and Iannazzo 2013). We also substantially extended the evaluation of numerical labeling experiments by adding multiple illustrations as well as quantitative results. This includes comparison to the additional reference method proposed in Rathke et al. (2014). Finally, we added discussion of feature locality and of variance in the reference segmentations used for training.

The remainder of this paper is organized as follows. The assignment flow approach is summarized in Sect. 2 and extended in Sect. 4 in order to take into account the order of layers as a global constraint. In Sect. 3, we consider the Riemannian manifold $\mathscr{P}_d$ of positive definite matrices as a suitable feature space for local OCT data descriptors. Various Riemannian metrics are discussed with regard to computational efficiency of clustering. The resulting features are subsequently compared to local features extracted by a convolutional network in Sect. 5. Performance measures for OCT segmentation will be reported for our novel approach and for two other state-of-the-art methods with available standalone software, that were evaluated in detail as summarized in Sect. 5. In Sect. 6, we shortly discuss the access to appropriate ground truth data and the impact of feature locality underlying our approach.

2 Assignment Flow

We summarize the assignment flow approach introduced by Åström et al. (2017) and refer to the recent survey (Schnörr 2020) for more background and a review of recent related work.

2.1 Overview

The assignment manifold $\mathcal{W}$ (16) is a product space of probability simplices. Hence each point $W \in \mathcal{W}$ is a collection of discrete probability vectors, one for each pixel, called assignment vectors. These vectors $W(t)$ evolve on $\mathcal{W}$ according to the assignment flow ODE (25). Due to the imposed Fisher–Rao geometry (12), $W(t)$ converges to an integral solution (Zern et al. 2020a): for $t \to \infty$, each $W_i(t)$ approaches an unit vector that encodes the class label j assigned to the data point f_i given at pixel $i \in I$.

Thus, assignment flows perform labelings as do discrete graphical models (Kappes et al. 2015). Yet, unlike the latter models, the assignment flow approach is *smooth* which enables efficient numerical inference (Zeilmann et al. 2020), parameter learning (Hühnerbein et al. 2021) and extensions to unsupervised and self-supervised scenarios (Zern et al. 2020b; Zisler et al. 2020).

Section 4 extends the assignment flow approach such that the natural ordering of labels due to retinal tissue layers is taken into account.

2.2 Assignment Manifold

Let $(\mathcal{F}, d_{\mathcal{F}})$ be a metric space and

$$\mathcal{F}_n = \{f_i \in \mathcal{F} : i \in I\}, \quad |I| = n \tag{1a}$$

given data. Assume that a predefined set of prototypes

$$\mathcal{F}_* = \{f_j^* \in \mathcal{F} : j \in J\}, \quad |J| = c \tag{1b}$$

is given. *Data labeling* denotes the assignments

$$j \to i, \quad f_j^* \to f_i \tag{2}$$

of a single prototype $f_j^* \in \mathcal{F}_*$ to each data point $f_i \in \mathcal{F}_n$. The set I is assumed to form the vertex set of an undirected graph $\mathcal{G} = (I, \mathcal{E})$ which defines a relation $\mathcal{E} \subset I \times I$ and neighborhoods

$$\mathcal{N}_i = \{k \in I : ik \in \mathcal{E}\} \cup \{i\}, \tag{3}$$

where ik is a shorthand for the unordered pair (edge) $(i, k) = (k, i)$. We require these neighborhoods to satisfy the symmetry relation

$$k \in \mathcal{N}_i \quad \Leftrightarrow \quad i \in \mathcal{N}_k, \qquad \forall i, k \in I. \tag{4}$$

The assignments (labeling) (2) are represented by matrices in the set

$$\mathcal{W}_* = \{W \in \{0, 1\}^{n \times c} : W \mathbb{1}_c = \mathbb{1}_n\} \tag{5}$$

with unit vectors W_i, $i \in I$, called *assignment vectors*, as row vectors. These assignment vectors are computed by numerically integrating the assignment flow below (25) in the following geometric setting. The integrality constraint of (5) is relaxed and vectors

$$W_i = (W_{i1}, \ldots, W_{ic})^\top \in \mathcal{S}, \quad i \in I, \tag{6}$$

that we still call *assignment vectors*, are considered on the elementary Riemannian manifold

$$(\mathcal{S}, g), \quad \mathcal{S} = \{p \in \Delta_c : p > 0\} \tag{7}$$

with the probability simplex

$$\Delta_c = \left\{ p \in \mathbb{R}_+^c : \sum_{i=1}^c = \langle \mathbb{1}, p \rangle = 1 \right\}, \tag{8}$$

the barycenter

$$\mathbb{1}_{\mathcal{S}} = \frac{1}{c} \mathbb{1}_c \in \mathcal{S}, \qquad \text{(barycenter)} \tag{9}$$

tangent space

$$T_0 = \{v \in \mathbb{R}^c : \langle \mathbb{1}_c, v \rangle = 0\} \tag{10}$$

and tangent bundle $T\mathcal{S} = \mathcal{S} \times T_0$, the orthogonal projection

$$\Pi_0 : \mathbb{R}^c \to T_0, \quad \Pi_0 = I - \mathbb{1}_{\mathcal{S}} \mathbb{1}^\top \tag{11}$$

and the Fisher–Rao metric

$$g_p(u, v) = \sum_{j \in J} \frac{u^j v^j}{p^j}, \quad p \in \mathcal{S}, \quad u, v \in T_0. \tag{12}$$

Based on the linear map

$$R_p : \mathbb{R}^c \to T_0, \quad R_p = \operatorname{Diag}(p) - pp^\top, \quad p \in \mathcal{S} \tag{13}$$

that satisfies

$$R_p = R_p \Pi_0 = \Pi_0 R_p, \tag{14}$$

exponential maps and their inverses are defined by

$$\operatorname{Exp} : \mathcal{S} \times T_0 \to \mathcal{S}, \ (p, v) \mapsto \operatorname{Exp}_p(v) = \frac{p e^{\frac{v}{p}}}{\langle p, e^{\frac{v}{p}} \rangle}, \tag{15a}$$

$$\operatorname{Exp}_p^{-1} : \mathcal{S} \to T_0, \qquad q \mapsto \operatorname{Exp}_p^{-1}(q) = R_p \log \frac{q}{p}, \tag{15b}$$

$$\exp_p : T_0 \to \mathcal{S}, \qquad \exp_p = \operatorname{Exp}_p \circ R_p, \tag{15c}$$

$$\exp_p^{-1}\colon \mathcal{S} \to T_0, \quad \exp_p^{-1}(q) = \Pi_0 \log \frac{q}{p} \tag{15d}$$

where multiplication, exponentials and logarithms apply componentwise. Applying the map $\exp_p$ to a vector in $\mathbb{R}^c = T_0 \oplus \mathbb{R}\mathbb{1}$ does not depend on the constant component of the argument, due to (14).

Remark 1 The map Exp corresponds to the e-connection of information geometry (Amari and Nagaoka 2000), rather than to the exponential map of the Riemannian connection. Accordingly, the affine geodesics (15a) are not length-minimizing. But they provide a close approximation (Åström et al. 2017, Prop. 3) and are more convenient for numerical computations.

The *assignment manifold* is defined as

$$(\mathcal{W}, g), \quad \mathcal{W} = \mathcal{S} \times \cdots \times \mathcal{S}. \quad (n = |I| \text{ factors}) \tag{16}$$

We identify $\mathcal{W}$ with the embedding into $\mathbb{R}^{n\times c}$

$$\mathcal{W} = \Big\{ W \in \mathbb{R}^{n\times c} \colon W\mathbb{1}_c = \mathbb{1}_n \text{ and } W_{ij} > 0 \text{ for all } i \in [n],\, j \in [c] \Big\}. \tag{17}$$

Thus, points $W \in \mathcal{W}$ are row-stochastic matrices $W \in \mathbb{R}^{n\times c}$ with row vectors $W_i \in \mathcal{S}$, $i \in I$ that represent the assignments (2) for every $i \in I$. We set

$$\mathcal{T}_0 := T_0 \times \cdots \times T_0 \quad (n = |I| \text{ factors}). \tag{18}$$

Due to (17), the tangent space $\mathcal{T}_0$ can be identified with

$$\mathcal{T}_0 = \{V \in \mathbb{R}^{n\times c} \colon V\mathbb{1}_c = 0\}. \tag{19}$$

Thus, $V_i \in T_0$ for all row vectors of $V \in \mathbb{R}^{n\times c}$ and $i \in I$. All mappings defined above factorize in a natural way and apply row-wise, e.g. $\mathrm{Exp}_W = (\exp_{W_1}, \ldots, \exp_{W_n})$ etc.

2.3 Assignment Flow

Based on (1a) and (1b), the distance vector field

$$D_{\mathcal{F};i} = \big(d_{\mathcal{F}}(f_i, f_1^*), \ldots, d_{\mathcal{F}}(f_i, f_c^*)\big)^\top, \quad i \in I \tag{20}$$

is well-defined. These vectors are collected as row vectors of the *distance matrix*

$$D_{\mathcal{F}} \in S_+^n, \tag{21}$$

where S_+^n denotes the set of symmetric and entrywise nonnegative matrices.

Remark 2 In this paper, we build upon two different types of features to determine vectors (20) which are serving as input before mapping the assembled matrix (21) onto the assignment manifold as explained below. Hereby, the first class of features access our model by calculating distance to prototypes (1) with metric introduced in section (Sect. 3.2) while the second feature class directly possess the form of (21) as argued in section (Sect. 5.2.3).

The *likelihood map* and the *likelihood vectors*, respectively, are defined for $i \in I$ as

$$L_i \colon \mathcal{S} \to \mathcal{S},$$
$$L_i(W_i) = \exp_{W_i}\Big(-\frac{1}{\rho} D_{\mathcal{F};i}\Big) = \frac{W_i e^{-\frac{1}{\rho} D_{\mathcal{F};i}}}{\langle W_i, e^{-\frac{1}{\rho} D_{\mathcal{F};i}} \rangle}, \tag{22}$$

where the scaling parameter $\rho > 0$ is used for normalizing the a-prior unknown scale of the components of $D_{\mathcal{F};i}$ that depends on the specific application at hand.

A key component of the assignment flow is the interaction of the likelihood vectors through *geometric* averaging within the local neighborhoods (3). Specifically, using weights

$$\omega_{ik} > 0 \quad \text{for all } k \in \mathcal{N}_i,\ i \in I \quad \text{with} \quad \sum_{k\in\mathcal{N}_i} w_{ik} = 1, \tag{23}$$

the *similarity map* and the *similarity vectors*, respectively, are defined for $i \in I$ as

$$S_i \colon \mathcal{W} \to \mathcal{S},$$
$$S_i(W) = \mathrm{Exp}_{W_i}\left(\sum_{k\in\mathcal{N}_i} w_{ik} \exp_{W_i}^{-1}\big(L_k(W_k)\big)\right). \tag{24}$$

If Exp_{W_i} were the exponential map of the Riemannian (Levi-Civita) connection, then the argument inside the brackets of the right-hand side would just be the negative Riemannian gradient with respect to W_i of center of mass objective function comprising the points L_k, $k \in \mathcal{N}_i$, i.e. the weighted sum of the squared Riemannian distances between W_i and L_k (Jost 2017, Lemma 6.9.4). In view of Remark 1, this interpretation is only approximately true mathematically, but still correct informally: $S_i(W)$ moves W_i towards the geometric mean of the likelihood vectors L_k, $k \in \mathcal{N}_i$. Since $\mathrm{Exp}_{W_i}(0) = W_i$, this mean precisely is W_i if the aforementioned gradient vanishes.

The *assignment flow* is induced on the assignment manifold $\mathcal{W}$ by the locally coupled system of nonlinear ODEs

$$\dot{W} = R_W S(W), \quad W(0) = \mathbb{1}_{\mathcal{W}}, \tag{25a}$$
$$\dot{W}_i = R_{W_i} S_i(W), \quad W_i(0) = \mathbb{1}_{\mathcal{S}}, \quad i \in I, \tag{25b}$$

where $\mathbb{1}_{\mathscr{W}} \in \mathscr{W}$ denotes the barycenter of the assignment manifold (16). The solution $W(t) \in \mathscr{W}$ is numerically computed by geometric integration (Zeilmann et al. 2020) and determines a labeling $W(T) \in \mathscr{W}_*$ for sufficiently large T after a trivial rounding operation. Convergence and stability of the assignment flow have been studied by Zern et al. (2020a).

3 OCT Data Representation by Covariance Desciptors

In this section, we work out the basic geometric notation for representation of OCT data by means of covariance descriptors (Tuzel et al. 2006). Specifically, the metric data space $(\mathscr{F}, d_{\mathscr{F}})$ underlying (1) will be identified with the Riemannian manifold $(\mathscr{P}_d, d_g)$ of positive definite matrices of dimension $d \times d$, with Riemannian metric g and Riemannian distance d_g as specified in Sect. 5. In particular regarding the computation of corresponding prototypes (1b), an important aspect concerns the trade-off between respecting the Riemannian distance d_g of the matrix manifold $\mathscr{P}_d$ and approximating surrogate distance functions, that enable to compute more efficiently Riemannian means of covariance descriptors while adopting their natural geometry. We review and discuss various choices in Sect. 3.2 after reviewing few required concepts of Riemannian geometry in Sect. 3.1.

3.1 The Manifold $\mathscr{P}_d$

We collect few concepts related to data $p \in \mathscr{M}$ taking values on a general Riemannian manifold $(\mathscr{M}, g)$ with Riemannian metric g; see, e.g., Lee (2013), Jost (2017) for background reading. Then we apply these concepts to the specific manifold $(\mathscr{P}_d, g)$ and the corresponding distance d_g, keeping the symbol g for the metric for simplicity. We refer to, e.g., Bhatia (2007, 2013), Pennec et al. (2006) and Moakher and Batchelor (2006) for further reading and to the references in Sect. 3.2.

Let $\gamma : [0, 1] \to \mathscr{M}$ a smooth curve connecting two points $p = \gamma(0)$ and $q = \gamma(1)$. The *Riemannian distance* between p and q is given by

$$d_g(p, q) = \min_{\gamma\,:\,\gamma(0)=p, \gamma(1)=q} L(\gamma) \tag{26a}$$

with

$$L(\gamma) = \int_0^1 \|\dot\gamma(t)\|_{\gamma(t)} \mathrm{d}t = \int_0^1 \sqrt{g_{\gamma(t)}\big(\dot\gamma(t), \dot\gamma(t)\big)}\mathrm{d}t. \tag{26b}$$

Assume the minimum of the right-hand side of (26a) is attained at $\overline{\gamma}$. Then the *exponential map* at p is defined on some neighborhood $V_p \subseteq T_p\mathscr{M}$ of 0 in the tangent space to $\mathscr{M}$ at p by

$$\begin{aligned} \exp_p \colon V_p \supseteq T_p\mathscr{M} &\to U_p \subseteq \mathscr{M}, \\ v &\mapsto \exp_p(v) := \overline{\gamma}(1). \end{aligned} \tag{27}$$

This mapping is a diffeomorphism of V_p and its inverse map $\exp_p^{-1} \colon U_p \to V_p$ exists on a corresponding open neighborhood U_p. Let $\mathscr{X}(\mathscr{M})$ denote the set of all smooth vector fields on $\mathscr{M}$, i.e. $X \in \mathscr{X}(\mathscr{M})$ evaluates to a tangent vector $X_p \in T_p\mathscr{M}$ smoothly depending on p. The set of all smooth covector fields (one-forms) is denoted by $\mathscr{X}^*(\mathscr{M})$, and $df(X)$ denotes the action of the differential $df \in \mathscr{X}^*(\mathscr{M})$ of a smooth function $f : \mathscr{M} \to \mathbb{R}$ on a vector field X. The *Riemannian gradient* of f is the vector field $\operatorname{grad} f \in \mathscr{X}(\mathscr{M})$ defined by

$$g(\operatorname{grad} f, X) = df(X) = Xf, \quad \forall X \in \mathscr{X}(\mathscr{M}). \tag{28}$$

We now focus on the following problem: Given a set of points $\{p_i\}_{i\in[N]} \subset \mathscr{M}$, compute the *weighted Riemannian mean* as minimizer of the objective function

$$\begin{aligned} &\overline{p} = \arg\min_{q\in\mathscr{M}} J(q), \quad J(q) = \sum_{i\in[N]} \omega_i d_g^2(q, p_i), \\ &\sum_{i\in[N]} \omega_i = 1, \qquad\qquad \omega_i > 0, \quad \text{for all } i. \end{aligned} \tag{29}$$

The Riemannian gradient of this objective function is given by Jost (2017, Lemma 6.9.4)

$$\operatorname{grad} J(p) = -\sum_{i\in[N]} \omega_i \exp_p^{-1}(p_i). \tag{30}$$

Hence the Riemannian mean $\overline{p}$ is determined by the optimality condition

$$\sum_{i\in[N]} \omega_i \exp_{\overline{p}}^{-1}(p_i) = 0. \tag{31}$$

A basic numerical method for computing $\overline{p}$ is the fixed point iteration

$$q_{(t+1)} = \exp_{q_{(t)}}\left(\sum_{i\in[N]} \omega_i \exp_{q_{(t)}}^{-1}(p_i)\right), \quad t = 1, 2, \ldots \tag{32}$$

that may converge for a suitable initialization $q_{(0)}$ to $\overline{p}$.

We now focus on the specific manifold $(\mathscr{P}_d, g)$

$$\mathscr{P}_d = \{S \in \mathbb{R}^{d\times d} \colon S = S^\top,\ S \text{ is positive definite}\} \tag{33}$$

with the tangent space

$$T_S\mathscr{P}_d = \{S \in \mathbb{R}^{d\times d} \colon S^\top = S\}, \tag{34}$$

equipped with the Riemannian metric

$$g_S(U,V) = \operatorname{tr}(S^{-1}US^{-1}V), \quad U,V \in T_S\mathscr{P}_d. \tag{35}$$

The Riemannian distance (26a) is given by

$$d_{\mathscr{P}_d}(S,T) = \left(\sum_{i\in[d]} \big(\log \lambda_i(S,T)\big)^2\right)^{1/2}, \tag{36}$$

whereas the exponential map (27) reads

$$\exp_S(U) = S^{\frac{1}{2}} \operatorname{expm}\left(S^{-\frac{1}{2}} U S^{-\frac{1}{2}}\right) S^{\frac{1}{2}}, \tag{37}$$

and $\operatorname{expm}(\cdot)$ denotes the matrix exponential. Finally, given a smooth objective function $J\colon \mathscr{P}_d \to \mathbb{R}$, the Riemannian gradient is given by

$$\operatorname{grad} J(S) = S\big(\partial J(S)\big)S \in T_S\mathscr{P}_d, \tag{38}$$

where the symmetric matrix $\partial J(S)$ denotes the Euclidean gradient of J at S. Since $\mathscr{P}_d$ is a simply connected, complete and nonpositively curved Riemannian manifold (Bridson and Häflinger 1999, Section 10), the exponential map (37) is globally defined and bijective, and the Riemannian mean always exists and is uniquely defined as minimizer of the objective function (29), after substituting the Riemannian distance (36).

3.2 Computing Prototypical Covariance Descriptors

In this section, we focus on the computational differential geometric framework required for extraction of prototypes (1b) as Riemannian means from a set of covariance descriptors assembled from OCT data. Application details are reported in Sect. 5. Particularly with regard to more efficient handling present volumetric data and to reduce the computational costs, a surrogate metrics and distances are reviewed in Sects. 3.2.2 and 3.2.3. Their qualitative comparison is reported in Sect. 5.

3.2.1 Computing Riemannian Means

Given a set of covariance descriptors

$$\mathscr{S}_N = \{(S_1,\omega_1),\ldots,(S_N,\omega_N)\} \subset \mathscr{P}_d \tag{39}$$

together with positive weights ω_i, we next focus on the solution of the problem (29) for specific geometry (33),

$$\overline{S} = \arg\min_{S\in\mathscr{P}_d} J(S;\mathscr{S}_N), \quad J(S;\mathscr{S}_N) = \sum_{i\in[N]} \omega_i d^2_{\mathscr{P}_d}(S,S_i), \tag{40}$$

with the distance $d_{\mathscr{P}_d}$ given by (36). From (37), we deduce

$$U = \exp_S^{-1} \circ \exp_S(U) = S^{\frac{1}{2}} \operatorname{logm}\left(S^{-\frac{1}{2}} \exp_S(U) S^{-\frac{1}{2}}\right) S^{\frac{1}{2}} \tag{41}$$

with the matrix logarithm $\operatorname{logm} = \operatorname{expm}^{-1}$ (Higham 2008, Section 11). As a result, optimality condition (31) reads

$$\sum_{i\in[N]} \omega_i \overline{S}^{\frac{1}{2}} \operatorname{logm}\left(\overline{S}^{-\frac{1}{2}} S_i \overline{S}^{-\frac{1}{2}}\right) \overline{S}^{\frac{1}{2}} = 0. \tag{42}$$

Applying the corresponding basic fixed iteration (32) has two drawbacks, however (Congedo et al. 2015): Convergence is not theoretically guaranteed and if the iteration converges, than at a linear rate only. Since each iterative step requires nontrivial numerical matrix decomposition that has to be applied multiple times to every voxel (vertex) of a 3D grid-graph, this results in an overall quite expensive approach, in particular when larger data sets are involved as is the case for highly resolved 3D OCT volumetric scans.

The following variant proposed by Bini and Iannazzo (2013) is guaranteed to converge at a *quadratic* rate assuming the matrices $\{S_1,\ldots,S_N\}$ to pairwise commute. Using the parametrization

$$S = LL^\top \tag{43}$$

corresponding to the Cholesky decomposition replacing the map of fixed point iteration (32) with its linearization leads to the following fixed point iteration

$$F_\tau(L;\mathscr{S}_N) = LL^\top - \tau \sum_{i\in[N]} \omega_i L^\top \operatorname{logm}(L^{-\top} S_i^{-1} L^{-1}) L, \tag{44}$$

with damping parameter $\tau > 0$. Comparing to (42) shows that the basic idea is to compute the Riemannian mean $\overline{S}$ as fixed point of the iteration

$$\overline{S} = \lim_{t\to\infty} S_{(t)}, \quad S_{(t+1)} = F(S_{(t)};\mathscr{S}_N). \tag{45}$$

Algorithm 1 provides a refined variant of this iteration including adaptive stepsize selection. See Congedo et al. (2015) for alternative algorithms that determine the Riemannian mean.

Algorithm 1: Fixed Point Iteration for Computing the Riemannian Matrix Mean.

Initialization

ϵ (termination threshold)

$t = 0, \quad S_{(0)} = LL^\top$, with $S_{(0)}$ solving (47).

$c_0 = \frac{\lambda_{\max}(S_{(0)})}{\lambda_{\min}(S_{(0)})}$, $\{\alpha_0, \beta_0\} = \left[\frac{\log(c_0)}{c_0 - 1}, c_0 \frac{\log(c_0)}{c_0 - 1}\right]$ (condition number and step size selection parameters)

$\tau_0 = \frac{2}{\alpha_0 + \beta_0}$

$S_{(1)} = F_\tau(L; \mathscr{S}_N)$ (iterative step)

$\epsilon_1 = \left\| \sum_{i \in [N]} \omega_i \operatorname{logm}(S_{(1)}^{\frac{1}{2}} S_i^{-1} S_{(1)}^{\frac{1}{2}}) \right\|_F, \quad t = 1$

while $\epsilon_t > \epsilon$ **do**

$S_{(t)} = LL^\top$

$c_t = \frac{\lambda_{\max}(S_{(t)})}{\lambda_{\min}(S_{(t)})}$

if $c_t = 1$ **then**

stop

$\{\alpha_t, \beta_t\} = \{\sum_{k=0}^{t} \frac{\log(c_k)}{c_k - 1}, c_k \frac{\log(c_k)}{c_k - 1}\}$

$\tau_t = \frac{2}{\alpha_t + \beta_t}$

$S_{(t+1)} = F_{\tau_t}(L; \mathscr{S}_N)$

$\epsilon_{t+1} := \left\| \sum_{i \in [N]} \omega_i \operatorname{logm}(S_{(t+1)}^{\frac{1}{2}} S_i^{-1} S_{(t+1)}^{\frac{1}{2}}) \right\|_F, \quad t \leftarrow t + 1$

3.2.2 Log-Euclidean Distance and Means

A computationally cheap approach was proposed by Arsigny et al. (2007) (among several other ones). Based on the operations

$$S_1 \odot S_2 = \operatorname{expm}\big(\operatorname{logm}(S_1) + \operatorname{logm}(S_2)\big), \tag{46a}$$

$$\lambda \cdot S = \operatorname{expm}\big(\lambda \operatorname{logm}(S)\big), \tag{46b}$$

the set $(\mathscr{P}_s, \odot, \cdot)$ becomes isomorphic to the vector space where $\odot$ plays the role of addition. Consequently, the mean of the data $\mathscr{S}_N$ given by (39) is defined analogous to the arithmetic mean by

$$\overline{S} = \operatorname{expm}\left(\sum_{i \in [N]} \omega_i \operatorname{logm}(S_i)\right). \tag{47}$$

While computing the mean is considerably cheaper than integrating the flow (38) using approximation Algorithm 1, the critical drawback of relying on (47) is not taking into account the (curved structure) of the manifold $\mathscr{P}_d$. Therefore, in the next section, we additionally consider another approximation of the Riemannian mean that better respects the underlying geometry but can still be evaluated more efficiently than the Riemannian mean of Sect. 3.2.1.

3.2.3 *S*-Divergence and Means

A general approach to the approximation of the objective function (29) is to replace the squared Riemannian $d_g^2(p, q)$ distance by a divergence function

$$D(p, q) \approx \frac{1}{2} d_g^2(p, q) \tag{48}$$

that satisfies

$$D(p, q) \geq 0 \quad \text{and} \quad D(p, q) = 0 \;\Leftrightarrow\; p = q, \tag{49a}$$

$$\partial_1^2 D(p, q) \succ 0, \quad \forall p \in \operatorname{dom} D(\cdot, q). \tag{49b}$$

We refer to, e.g., Bauschke and Borwein (1997) and Censor and Zenios (1997) for a complete definition. Property (49b) says that, for any feasible p, the Hessian with respect to the first argument is positive definite. In fact, suitable divergence functions D recover in this way locally the metric tensor of the underlying manifold $\mathscr{M}$, in order to qualify as a surrogate for the squared Riemannian distance (48).

For the present case $\mathscr{M} = \mathscr{P}_d$ of interest, Sra (2016) proposed the divergence function, called *Stein divergence* and is given for $S, S_1, S_2 \in \mathscr{P}_d$ as

$$D_s(S_1, S_2) = \log\det\left(\frac{S_1 + S_2}{2}\right) - \frac{1}{2}\log\det(S_1 S_2). \tag{50}$$

Regarding the task of evaluating the Riemannian distance (36), which is required for the second term of problem (40) for subsequential extraction of prototypes (1b) in Sect. 5, while avoiding to solve the numerically involved numerical generalized eigenvalue problem, we replace (40) by

$$\overline{S} = \arg\min_{S \in \mathscr{P}_d} J_s(S; \mathscr{S}_N), \quad J_s(S; \mathscr{S}_N) = \sum_{i \in [N]} \omega_i D_s(S, S_i). \tag{51}$$

The resulting Riemannian gradient flow reads

$$\dot{S} = -\operatorname{grad} J_s(S; \mathscr{S}_N) \overset{(38)}{=} -S \partial J(S; \mathscr{S}_N) S \tag{52a}$$

$$= -\frac{1}{2}\big(SR(S; \mathscr{S}_N)S - S\big), \tag{52b}$$

with

$$R(S; \mathscr{S}_N) = \sum_{i \in [N]} \omega_i \left(\frac{S + S_i}{2}\right)^{-1}. \tag{53}$$

Discretizing the flow using the geometric explicit Euler scheme with step size h,

$$S_{(t+1)} = \exp_{S_{(t)}}\big(-h \operatorname{grad} J_s(S_{(t)}; \mathscr{S}_N)\big) \tag{54a}$$

$$\overset{(37)}{=} S_{(t)}^{\frac{1}{2}} \operatorname{expm}\left(\frac{h}{2}\left(I - S_{(t)}^{\frac{1}{2}} R(S_{(t)}; \mathscr{S}_N) S_{(t)}^{\frac{1}{2}}\right)\right) S_{(t)}^{\frac{1}{2}} \tag{54b}$$

and using the log-Euclidean mean (47) as initial point $S_{(0)}$, defines Algorithm 2 as listed below.

Algorithm 2: Computing the Geometric Matrix Mean Based on the S-divergence.

Initialization
ϵ (termination threshold)
$t = 0, \quad S_{(0)}$ solves (47)
$\epsilon_0 > \epsilon$ (any value ϵ_0)
while $\epsilon_t > \epsilon$ **do**
$LL^\top = S_{(t)}$
$L_i L_i^\top = \frac{S_{(t)} + S_i}{2}$ for $i \in [N]$
$U = I - S_{(t)}^{\frac{1}{2}} \big(\sum_{i \in [N]} \omega_i (L_i L_i^\top)^{-1} \big) S_{(t)}^{\frac{1}{2}}$
$S_{(t+1)} = S_{(t)}^{\frac{1}{2}} \operatorname{expm}(\frac{h}{2} U) S_{(t)}^{\frac{1}{2}}$
$\epsilon_{t+1} := \|U\|_F, \quad t \leftarrow t + 1$

4 Ordered Layer Segmentation

In this section, we work out an extension of the assignment flow (Sect. 2) which is able to respect the order of cell layers as a global constraint while remaining in the same smooth geometric setting. In particular, existing schemes for numerical integration still apply to the novel variant.

4.1 Ordering Constraint

With regard to segmenting OCT data volumes, the order of cell layers is crucial prior knowledge. In this paper we focus on segmentation of the following 11 retina layers: retinal nerve fiber layer (RNFL), ganglion cell layer (GCL), inner nuclear layer (INL), outer plexiform layer (OPL), outer nuclear layer (ONL), two photoreceptor layers (PR1, PR2) separated by the external limiting membrane (ELM), Choriocapillaris (CC) and the retinal pigment epithelium (RPE) together with the choroid section (CS). Figure 3 also contains positions for the internal limiting membrane (ILM) and Bruch's membrane Membrane (BM).

To incorporate this knowledge into the geometric setting of Sect. 2, we require a smooth notion of ordering which allows to compare two probability distributions. In the following, we assume prototypes $f_j^* \in \mathscr{F}$, $j \in [n]$ in some feature space $\mathscr{F}$ to be indexed such that ascending label indices reflect the physiological order of cell layers.

Definition 1 (*Ordered Assignment Vectors*) A pair of voxel assignments $(w_i, w_j) \in \mathscr{S}^2$, $i < j$ within a single A-scan is called *ordered*, if $w_j - w_i \in K = \{By \colon y \in \mathbb{R}^c_+\}$ with the matrix

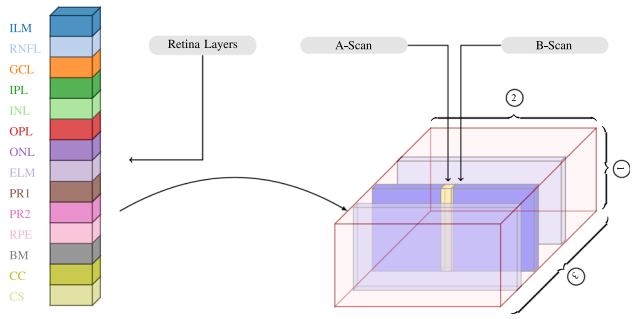


Fig. 3 OCT volume acquisition: ① is the A-scan axis (single A-scan is marked yellow). Multiple A-scans taken in rapid succession along axis ② form a two-dimensional B-scan (single B-scan is marked blue). The complete OCT volume is formed by repeating this procedure along axis ③. A list of retina layers that we expect to find in every A-scan is shown on the left (Color figure online)

$$B = \begin{pmatrix} -1 & & & \\ 1 & -1 & & \\ & 1 & \ddots & \\ & & \ddots & -1 \\ & & & 1 \quad -1 \end{pmatrix} \in \mathbb{R}^{c \times c}. \tag{55}$$

This new continuous ordering of probability distributions is consistent with discrete ordering of layer indices in the following way.

Lemma 1 *Let $w_i = e_{l_1}$, $w_j = e_{l_2}$, $l_1, l_2 \in [c]$ denote two integral voxel assignments. Then $w_j - w_i \in K$ if and only if $l_1 \leq l_2$.*

Proof B is regular with inverse

$$B^{-1} = -Q, \quad Q_{i,j} = \begin{cases} 1 & \text{if } i \geq j \\ 0 & \text{else} \end{cases} \tag{56}$$

and $w_j - w_i \in K \Leftrightarrow B^{-1}(w_j - w_i) \in \mathbb{R}^c_+$. It holds

$$B^{-1}(w_j - w_i) = Qe_{l_1} - Qe_{l_2} = \sum_{k=l_1}^{c} e_k - \sum_{k=l_2}^{c} e_k \tag{57}$$

such that $B^{-1}(w_j - w_i)$ has nonnegative entries exactly if $l_1 \leq l_2$. □

The continuous notion of order preservation put forward in Definition 1 can be interpreted in terms of a related discrete graphical model. Consider a graph consisting of two nodes connected by a single edge. The order constrained image labeling problem on this graph can be written as the integer linear program

$$\min_{W \in \{0,1\}^{2 \times c}, M \in \Pi(w_i, w_j)} \langle W, D \rangle + \theta \langle Q - \mathbb{I}, M \rangle \tag{58}$$

where $\Pi(w_i, w_j)$ denotes the set of coupling measures for marginals w_i, w_j and $\theta \gg 0$ is a penalty associated with violation of the ordering constraint. By taking the limit $\theta \to \infty$ we find the more tightly constrained problem

$$\min_{W \in \{0,1\}^{2\times c}, M \in \Pi(w_i, w_j)} \langle W, D \rangle \quad \text{s.t. } \langle Q - \mathbb{I}, M \rangle = 0. \tag{59}$$

Its feasible set has an informative relation to Definition 1 examined in Proposition 1.

Lemma 2 *Let $M \in \mathbb{R}^{c\times c}$ be an upper triangular matrix with non-negative entries above the diagonal and non-negative marginals*

$$M \mathbb{1}_c \geq 0, \quad M^\top \mathbb{1}_c \geq 0. \tag{60}$$

Then there exists a modified matrix M^1 with the same properties such that $M^1 \geq 0$.

Proof Equation (60) directly implies $M_{11} \geq 0$ and $M_{cc} \geq 0$ because M is upper triangular. For row indices $l \neq m$ and column indices $q \neq r$, define the matrix $O^{lm,qr}$ with

$$O^{lm,qr}_{ij} = \begin{cases} -1 & \text{if } (i,j) = (l,q) \vee (i,j) = (m,r) \\ 1 & \text{if } (i,j) = (l,r) \vee (i,j) = (m,q) \\ 0 & \text{else} \end{cases}. \tag{61}$$

Then $O^{lm,qr}\mathbb{1} = (O^{lm,qr})^\top \mathbb{1} = 0$. Adding a matrix $O^{lm,qr}$ to M does therefore not change its marginals, but it redistributes mass from the positions (l,q) and (m,r) to the positions (l,r) and (m,q). Due to (60), it is possible to choose scalars $\alpha^k_{lr} \geq 0$ such that

$$M + \sum_{2\leq k\leq c-1} \sum_{\substack{l<k \\ r>k}} \alpha^k_{lr} O^{lk,kr} \geq 0. \tag{62}$$

□

Proposition 1 *A pair of voxel assignments $(w_i, w_j) \in \mathscr{S}^2$ within an single A-scan is ordered if and only if the set*

$$\Pi(w_i, w_j) \cap \{M \in \mathbb{R}^{c\times c} \colon \langle Q - \mathbb{I}, M \rangle = 0\} \tag{63}$$

is not empty.

Proof See Appendix A. □

Proposition 1 shows that transportation plans between ordered voxel assignments w_i and w_j exist which do not move mass from w_{i,l_1} to w_{j,l_2} if $l_1 > l_2$. This characterizes order preservation for non-integral assignments as put forward in Definition 1.

4.2 Ordered Assignment Flow

Likelihoods as defined in (22) emerge by lifting $-\frac{1}{\rho} D_{\mathscr{F}}$ regarded as Euclidean gradient of $-\frac{1}{\rho}\langle D_{\mathscr{F}}, W\rangle$ to the assignment manifold. It is our goal to encode order preservation into a generalized likelihood matrix $L_{\text{ord}}(W)$. To this end, consider the assignment matrix $W \in \mathscr{S}^N$ for a single A-scan consisting of N voxels. We define the related matrix $Y(W) \in \mathbb{R}^{N(N-1)\times c}$ with rows indexed by pairs $(i,j) \in [N]^2$, $i \neq j$ in fixed but arbitrary order. Using the matrix Q defined by (56), let the rows of Y be given by

$$Y_{(i,j)}(W) = \begin{cases} Q(w_j - w_i) & \text{if } i > j \\ Q(w_i - w_j) & \text{if } i < j \end{cases}. \tag{64}$$

By construction, an A-scan assignment W is ordered exactly if all entries of the corresponding $Y(W)$ are nonnegative. This enables to express the ordering constraint on a single A-scan in terms of the energy objective

$$E_{\text{ord}}(W) = \sum_{(i,j)\in[N]^2,\, i\neq j} \phi(Y_{(i,j)}(W)). \tag{65}$$

where $\phi \colon \mathbb{R}^c \to \mathbb{R}$ denotes a smooth approximation of $\delta_{\mathbb{R}^c_+}$. In our numerical experiments, we choose

$$\phi(y) = \left\langle \gamma \exp\left(-\frac{1}{\gamma} y\right), \mathbb{1} \right\rangle \tag{66}$$

with a constant $\gamma > 0$. Suppose a full OCT volume assignment matrix $W \in \mathscr{W}$ is given and denote the set of submatrices for each A-scan by $C(W)$. Then order preserving assignments consistent with given distance data $D_{\mathscr{F}}$ in the feature space $\mathscr{F}$ are found by minimizing the energy objective

$$E(W) = \langle D_{\mathscr{F}}, W \rangle + \sum_{W_A \in C(W)} E_{\text{ord}}(W_A). \tag{67}$$

We consequently define the generalized likelihood map

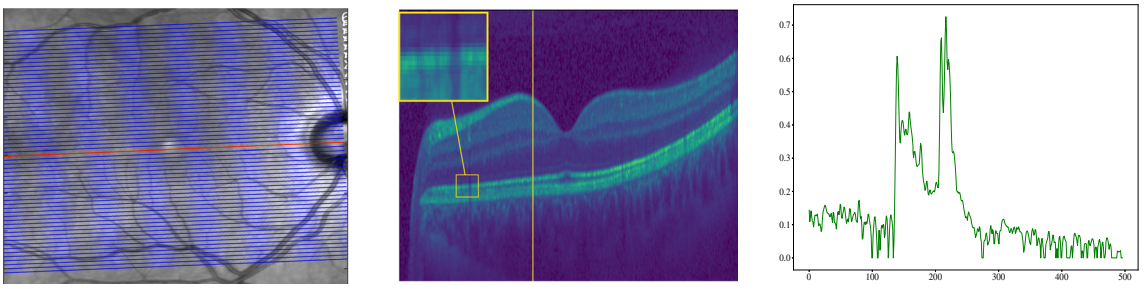
$$\begin{aligned} L_{\text{ord}}(W) &= \exp_W(-\nabla E(W)) \\ &= \exp_W\left(-\frac{1}{\rho} D_{\mathscr{F}} - \sum_{W_A \in C(W)} \nabla E_{\text{ord}}(W_A)\right) \end{aligned} \tag{68}$$

and specify a corresponding assignment flow variant.

Definition 2 (*Ordered Assignment Flow*) The dynamical system

$$\dot{W} = R_W S(L_{\text{ord}}(W)), \quad W(0) = \mathbb{1}_{\mathscr{W}} \tag{69}$$

evolving on $\mathscr{W}$ is called the *ordered assignment flow*.



(a) (b) (c)

Fig. 4 *Left* En-face view on the volumetric OCT data superimposed by parallel blue lines which represent the location of 61 B-scans within the volume. The red line indicates the position of a B-scan shown in the center image. *Center* The enlarged view on a B-scan depicts typical artifacts such as shadow regions and speckle noise. *Right* The gray value intensity of a single vertical A-scan located near the Fovea region. This A-scan is highlighted by a yellow line in the enlarged B-scan (center image). Noisy intensity variations along the A-scan indicate the difficulty of automatically extracting retinal tissue boundary positions (Color figure online)

By applying known numerical schemes (Zeilmann et al. 2020) for approximately integrating the flow (69), we find a class of discrete-time image labeling algorithms which respect the physiological cell layer ordering in OCT data. In Sect. 5, we benchmark the simplest instance of this class, emerging from the choice of geometric Euler integration.

5 Experimental Results

5.1 Data, Competing Approaches, Performance Measures

5.1.1 OCT-Data

In the following sections, after introducing key terminology in volumetric OCT data we describe experiments performed on a set of OCT volumes depicting the intensity of light reflection in chorioretinal tissues centered around the fovea. The scans were obtained using a spectral domain OCT device (Heidelberg Engineering, Germany) for multiple patients at a variety of resolutions by averaging various registered B-scans which share the same location in order to reduce speckle noise. This is representative of the fact that different resolutions may be desirable in clinical settings at the preference of medical practitioners. In the following, we always assume an OCT volume in question to consist of N_B B-scans, each comprising N_A A-scans with N voxels and use the term surface to refer to the set of voxels located at the interface of two retina layers. See Fig. 3 for a schematic illustration of the data acquisition process.

In the present work, we use a private dataset of 3D OCT volume scans provided by Heidelberg Engineering GmbH which we split into 82 volumes for training and 8 volumes for testing. In particular, the test set contains scans from multiple different patients without any observable pathological retina changes. See Appendix C for a detailed list of volume sizes and resolutions along each axis.

Figure 4 demonstrates the typical organization of a 3D-OCT volume acquired by scanning healthy human retina using an OCT device. B-Scans are indicated as blue lines placed in the Fundus image on the left. The particular B-Scan marked in red is depicted in the middle of Fig. 4. This illustrates the typical artifacts and corrupted layer intensities of the OCT volume. The right plot depicts the noisy signal along an A-scan indicated by a yellow vertical line which underpins the difficulty of segmenting the underlying data sets.

5.1.2 Reference Methods

To assess the segmentation performance of our proposed approach we compare ourselves to state of the art retina segmentation methods presented in Rathke et al. (2014) and Kang et al. (2006) which are applicable for both healthy and pathological patient data. In particular, we prefer these reference methods over Dufour et al. (2013), Song et al. (2013) and Garvin et al. (2009) because available implementations of the latter are limited to the segmentation of up to 9 retina layers. For both reference methods, we use the software implementation of their authors without any additional tuning or retraining.

IOWA Reference Algorithm A well-known graph-based approach to segmentation of macular volume data was developed by the Retinal Image Analysis Laboratory at the Iowa Institute for Biomedical Imaging (Kang et al. 2006; Abràmoff et al. 2010; Garvin et al. 2009). The problem of localizing cell layer boundaries in 3D OCT volumes is posed and ultimately transformed into a minimum st-cut problem on a non-trivially constructed graph G. To this end, a distance

tensor $D_k \in \mathbb{R}^{N_B \times N_A \times N}$ is formed in a feature extraction step for each boundary $k \in [c-1]$. This encodes $c-1$ separate binary segmentation problems on a geometric graph G_k spanning the volume. In each instance, voxels are to be classified as either belonging to boundary k or not belonging to boundary k. By utilizing a (directed) neighborhood structure on each G_k, smoothness constraints are introduced and regulated via user-specified stiffness parameters. To model interactions between different boundaries, the graphs G_k are combined to a global graph G, introducing additional edges between them. The latter set up constraints on the distance between consecutive boundaries within each A-scan which can be used to enforce physiological ordering of cell layers. On G, the problem of optimal boundary localization takes the form of minimal closed set construction which is in turn transformed into a minimum st-cut problem for which standard methods exist. Their standalone software is freely available for research purposes.[1]

Probabilistic Model Rathke et al. (2014) proposed a graph-based probabilistic approach for segmenting OCT volumes for given data y by leveraging the Bayesian ansatz

$$p(y, s, b) = p(y|s)p(s|b)p(b). \tag{70}$$

Here, the tensor $b \in \mathbb{R}^{N_B \times N_A \times (c-1)}$ contains real-valued boundary positions between retina layers and s denotes discrete (voxel-wise) segmentation. The appearance terms $p(y|s)$, $p(s|b)$ and $p(b)$ represent data likelihood, Markov random field regularizer and global shape prior respectively. In order to approximate the desired posterior

$$p(b, s|y) = \frac{p(y|s)p(s|b)p(b)}{p(y)}, \tag{71}$$

a variational inference strategy is employed. This aims to find a tractable distribution q decoupled into

$$q(b, s) = q_b(b)q_s(s) \tag{72}$$

which is close to $p(b, s|y)$ in terms of the relative entropy $\mathrm{KL}(q \mid p)$. The shape prior $p(b)$ is learned offline by maximum likelihood estimation in the space of normal distributions using a low-rank approximation of the involved covariance matrix. Ordering constraints

$$1 \le s_{1,ij} \le s_{2,ij} \le \cdots \le s_{c-1,ij}, \quad ij \in [N_B] \times [N_A] \tag{73}$$

are enforced for the discrete segmentation s and are not enforced for the continuous boundaries b. This is in contrast to the proposed model which integrates the ordering of retina layers by adding a cost function (63) penalizing the overall deviation of soft assignments during numerical integration of (25) from the subspace of probability distributions satisfying (1). The method comes along with a standalone software which is freely available.[2]

5.1.3 Performance Measures

We will evaluate the computed segmentations by their direct comparison with manual annotations regarded as gold standard which were realized by a medical expert. Respective metrics are suitable for segmentation tasks that involve multiple tissue types (Crum et al. 2006). Specifically, we report the mean DICE similarity coefficient (Dice 1945) for each segmented cell layer.

Definition 3 (*DICE*) Given two sets A, B the *DICE similarity coefficient* is defined as

$$\mathrm{DSC}(A, B) := \frac{2|A \cap B|}{|A| + |B|} = \frac{2TP}{2TP + FP + FN} \in [0, 1], \tag{74}$$

where $\{TP, FN, FP\}$ denotes the number of *true positives*, *false negatives* and *false positives* respectively.

The DICE similarity coefficient quantifies the region agreement between computed segmentation results and manually labeled OCT volumes which serve as ground truth. High similarity index $\mathrm{DSC}(A, B) \approx 1$ indicates large relative overlap between the sets A and B. This metric is well suited for average performance evaluation and appears frequently in the literature (e.g. Chiu et al. 2015; Yazdanpanah et al. 2011; Novosel et al. 2017). It is closely related to the positively correlated Jaccard similarity measure (Jaccard 1908) which in contrast to (74) is more strongly influenced by worst case performance.

In addition, we report the mean absolute error (MAE) of computed layer boundaries used in Rathke et al. (2014) and Garvin et al. (2009) to make our results more directly comparable to these references.

Definition 4 (*Mean Absolute Error*) For a single A-scan indexed by $ij \in [N_B] \times [N_A]$, let $e_{ij} := |g_{ij} - p_{ij}|$ denote the absolute difference between a layer boundary position g_{ij} in the gold standard segmentation and a predicted layer boundary p_{ij}. The mean absolute error (MAE) is defined as the mean value

$$\mathrm{MAE}(g, p) = \frac{1}{N_B N_A} \sum_{ij \in [N_B] \times [N_A]} e_i. \tag{75}$$

[1] see https://www.iibi.uiowa.edu/oct-reference.

[2] https://github.com/FabianRathke/octSegmentation.

5.2 Feature Extraction

5.2.1 Region Covariance Descriptors

To apply the geometric framework proposed in Sect. 3 we next introduce the *region covariance descriptors* (Tuzel et al. 2006) which have been widely applied in computer vision and medical imaging, see e.g. Cherian and Sra (2016), Turaga and Srivastava (2016), Depeursinge et al. (2014) and Sirinukunwattana et al. (2015). We model the raw intensity data for a given OCT volume by a mapping $I : \mathscr{D} \to \mathbb{R}_+$ where $\mathscr{D} \subset \mathbb{R}^3$ is the underlying spatial domain. To each voxel $v \in \mathscr{D}$, we associate the local feature vector $f : \mathscr{D} \to \mathbb{R}^{10}$,

$$f : \mathscr{D} \to \mathbb{R}^{10} \tag{76}$$

$$v \mapsto (I(v), \nabla_x I(v), \nabla_y I(v), \nabla_z I(v), \sqrt{2}\nabla_{xy} I(v), \ldots, \nabla_{zz} I(v))^\top. \tag{77}$$

assembled from the intensity $I(v)$ as well as first- and second-order responses of derivative filters capturing information from larger scales following (Hashimoto and Sklansky 1987). To improve the segmentation accuracy we combine the derivative filter responses from various scales in an computationally efficient way we first normalize the derivatives of the input volume $I(v)$ at every scale σ_s by convolution each dimension with a 1D window:

$$\nabla_x \tilde{I}_{\sigma_s}(v) = \sigma_s^2 \frac{\partial}{\partial x} \tilde{G}(v, \sigma_s) \tag{78}$$

where $\tilde{G}(v, \sigma_s)$ is an approximation to a Gaussian window $\big(G(v, \sigma_s) * I\big)(v)$ at scale σ_s as in detail described in Hashimoto and Sklansky (1987). Subsequently we follow the idea presented by Lindeberg (2004) by taking local maxima over scales

$$\nabla_x \tilde{I}(v) = \max_{\sigma_s} \nabla_x \tilde{I}_{\sigma_s}(v), \tag{79}$$

which are serving for the mapping (76).

By introducing a suitable geometric graph spanning $\mathscr{D}$, we can associate a neighborhood $\mathscr{N}_i$ of fixed size with each voxel $i \in [n]$ as in (24). For each neighborhood, we define the regularized *region covariance descriptor*

$$S_i := \sum_{j \in \mathscr{N}_i} \theta_{ij} (f_j - \overline{f_i})(f_j - \overline{f_i})^T + \epsilon I, \quad \overline{f_i} = \sum_{k \in \mathscr{N}_i} \theta_{ik} f_k, \tag{80}$$

as a weighted empirical covariance matrix with respect to feature vectors f_j. The small value $1 \gg \epsilon > 0$ acts as a regularization parameter enforcing positive definiteness of S_i. Diagonal entries of each covariance matrix C_i are empirical variances of feature channels in (76) while the off-diagonal entries represent empirical correlations within the region $\mathscr{N}_i$.

5.2.2 Prototypes on $\mathscr{P}^d$

In view of the assignment flow framework introduced in Sect. 2, we interpret region covariance descriptors (80) as data points in the metric space $\mathscr{P}^d$ of symmetric positive definite matrices and model each retina tissue indexed by $l \in [c]$ with a random variable S_l taking values in $\mathscr{P}^d$. Suppose we draw N_l samples $\{S_l^k\}_{k=1}^{N_l}$ from the distribution of S_l. The most basic way to apply assignment flows to data in $\mathscr{P}^d$ is based on computing a prototypical element of $\mathscr{P}^d$ for each tissue layer, e.g. the Riemannian center of mass of $\{S_l^k\}_{k=1}^{N_l}$. This corresponds to directly choosing $\mathscr{P}^d$ as feature space $\mathscr{F}$ in (1a). We find that superior empirical results are achieved by considering a dictionary of $K_l > 1$ prototypical elements for each layer $l \in [c]$. This entails partitioning the samples $\{S_l^k\}_{k=1}^{N_l}$ into K_l disjoint subsets $\hat{S}_l^j \subseteq \{S_l^k\}_{k=1}^{N_l}$, $j \in [K_l]$ with representatives $\tilde{S}_l^j$ determined offline.

To find a set of representatives which captures the structure of the data, we minimize expected loss measured by the Stein divergence (50) leading to the K-means like functional

$$\mathbb{E}_{p_l}(\tilde{S}_l) = \sum_{j=1}^{K_l} p(j) \sum_{S_l^i \in \hat{S}_l^j} \frac{p(i|j)}{p(j)} D_S(S_l^i, \tilde{S}_l^j), \qquad p(i,j) = \frac{1}{N_l}, \; p_l(j) = \frac{N_j}{N_l}. \tag{81}$$

A hard partitioning is achieved by applying Lloyd's algorithm in conjunction with Algorithm 2 for mean retrieval. We additionally employ the more common soft K-means like approach for determining prototypes by employing the mixture exponential family model based on Stein divergence to given data

$$p(S_l^i, \Gamma_l) = \sum_{j=1}^{K} \pi_l^j \, p(S_l^i, \tilde{S}_l^j)), \tag{82}$$

where the parameters

$$\Gamma_l = \{(\pi_l^j\}_{j=1}^K, \{\tilde{S}_l^j\}_{j=1}^K), \quad (\pi_l^1, \ldots, \pi_l^{|J|}) \in S \tag{83}$$

have to be adjusted to given data. The prototypes are recovered as mean parameters $S_l^{j,T}$ though an iterative process commonly refered to as *expectation maximation* (EM) defined by alternation of the following iterations

$$p_l(j|S_l^i, \Gamma_l^t) = \frac{\pi_l^{(j,t)} e^{-D_S(S_l^i, \tilde{S}_l^{(j,t)})}}{\sum_{k=1} \pi_l^{(k,t)} e^{-D_S(S_l^i, \tilde{S}_l^{(k,t)})}}, \tag{84}$$

(***Expectation step***)

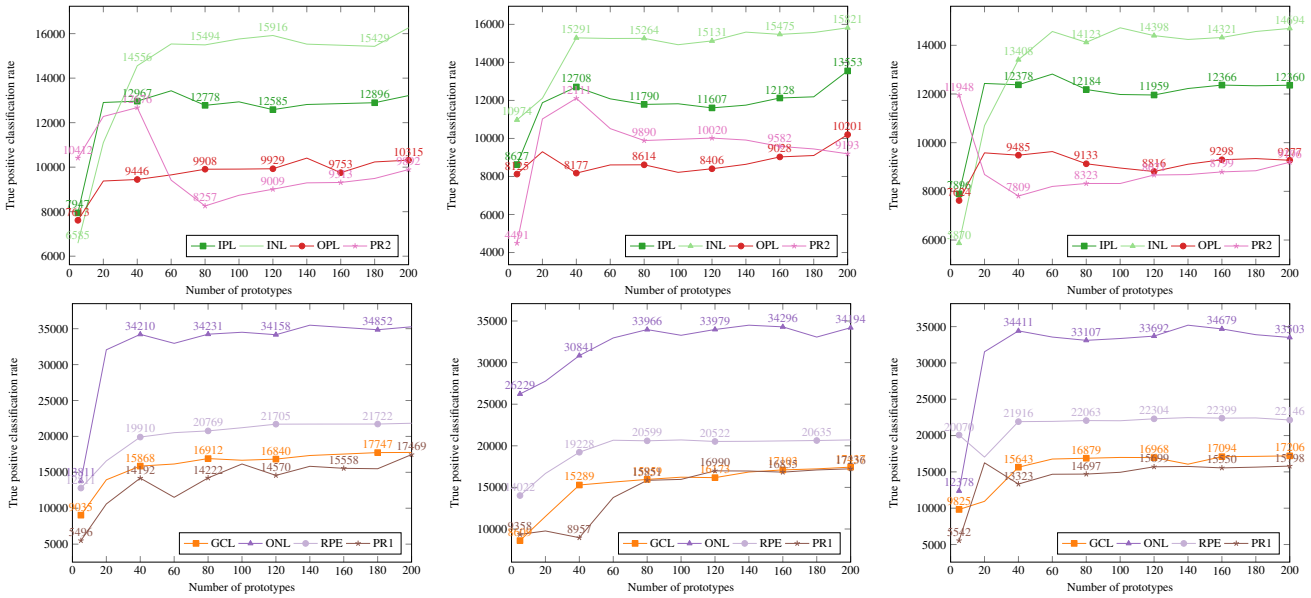


Fig. 5 *Top* Metric classification evaluated on thin layers (IPL, INL, OPL, PR2). *Bottom* Analogous metric evaluation for (GCL, ONL, PR1, RPE). *From left to right* The number of true outcomes after direct comparison with ground truth, for the choice of the exact Riemannian geometry of $\mathscr{P}_d$, Stein divergence and Log-Euclidean distance for geometric mean computation. The results of first two columns indicate higher detection performance while respecting the Riemannian geometry of a curved manifold. Enlarging the set of prototypical covariance descriptors leads to increased matching accuracy which is in contrast to the observed flattening of matching curves when using the Log-Euclidean distance

followed by updating the marginals at each time step up to final time T

$$\pi_l^{(j,t+1)} = \sum_{i=1}^{N_j} p_l(j|S_l^i, \Gamma_l^t)\tilde{S}^{j,t} \tag{85}$$

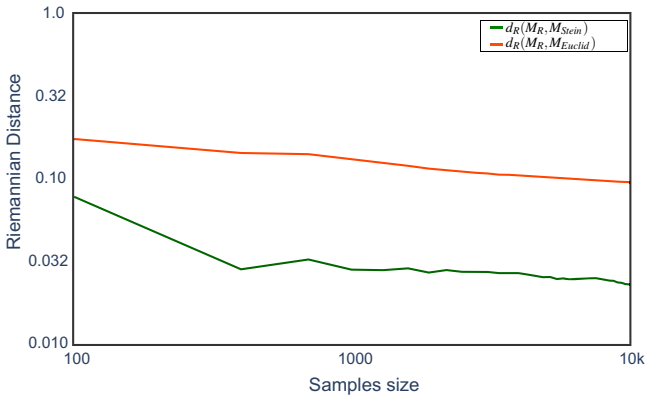
$$\tilde{S}_l^{j,t+1} = \operatorname{argmin}_{S\in\mathscr{P}_d}\left(\sum_{i=1}^{n} p(j|\Gamma_i^t) D_S(S_l^i, S)\right). \quad (\textbf{\textit{Maximization step}}) \tag{86}$$

The decision to approximate the Riemannian metric on $\mathscr{P}_d$ by the Stein divergence (50) can be backed up empirically. To this end, we randomly select descriptors (80) representing the nerve fibre layer in real-world OCT data and compute their Riemannian mean as well as their mean w.r.t. the Log-Euclidian metric (46) and Stein divergence (50). Figure 6 illustrates that Stein divergence approximates the full Riemannian metric more precisely than the Log-Euclidian metric while still achieving a significant reduction in computational effort. Furthermore to evaluate the classification we extracted a dictionary of 200 prototypes for representing each retina tissue for different choice of metric and subsequently evaluated the resulting segmentation accuracy by assigning each voxel to a class containing the prototype with smallest distance using a cropped OCT Volume of size $138 \times 100 \times 40$ taken from the testing set.

Figure 5 visualizes the correct classification matches for retina layers ordered by color according to Fig. 3. In particular, we inspect a notable gain of correct matches while respecting the Riemannian geometry (first column) as opposed to Log-Euclidean setting (third column). Regarding the approximation of (36) by (50), we are observing more effective detection of outer photoreceptor layer (PR1), inner nuclear layer (INL) and retinal pigment epithelium (RPE). Furthermore, taking a closer look at (OPL) and (ONL) we note a typical tradeoff between the number of prototypes and detection performance indicating superior retina to voxel allocation by applying (46), whereas the surrogate divergence metric (50) has the tendency to improve the accuracy while increasing the size of evaluated prototypes in contrast to flattening curves when relying on (47).

This illustrates a tradeoff between computational effort and labeling performance, cf. Fig. 6. Note that prototypes are computed offline, making runtime performance less relevant to medical practitioners. However, building a distance matrix involves computing $n \sum_{l\in[c]} K_l$ Riemannian distances resp. Stein divergences to prototypes. This still leads to a large difference in (online) runtime since evaluation of the Riemannian distance (36) involves generalized eigendecomposition while less costly Cholesky decomposition suffices to evaluate the Stein divergence (50).

Summarizing the discussed results concerning the application of Algorithms 1 and 2, we point out that respecting the Riemannian geometry leads to superior labeling results providing more descriptive prototypes (Figs. 5, 6).



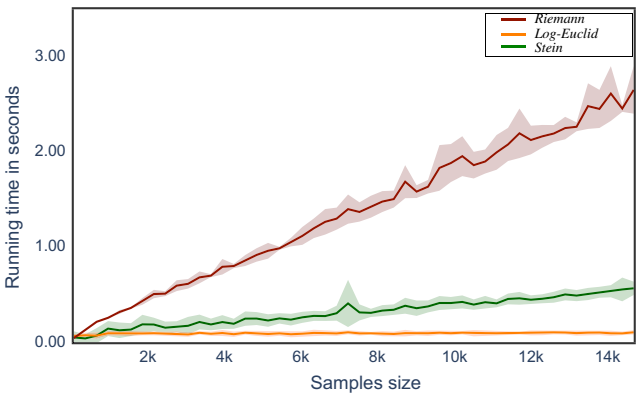


Fig. 6 *Left* Deviation of the geometric means computed using the Log-Euclidian metric and Stein divergence, respectively, from the true Riemannian mean. *Right* Runtime for geometric mean computation using the different metrics. All evaluations were performed on a randomly chosen subset of covariance descriptors representing the retinal nerve fibre layer in a real-world OCT scan. Both graphics clearly highlight the advantages of using Stein the divergence in terms of approximation accuracy and efficient numerical computation

5.2.3 CNN Features

In addition to the covariance features in Sect. 5.2.1, we compare a second approach to local feature extraction based on a convolutional neural network architecture. For each node $i \in [n]$, we trained the network to directly predict the correct class in $[c]$ using raw intensity values in $\mathscr{N}_i$ as input. As output, we find a score for each layer which can directly be transformed into a distance vector suitable as input to the ordered assignment flow (69) via (68). The specific network used in our experiments has a ResNet architecture comprising four residually connected blocks of 3D convolutions and ReLU activation. Model size was hand-tuned for different sizes of input neighborhoods, adjusting the number of convolutions per block as well as corresponding channel dimensions. Details of the employed architecture are listed in Appendix B. In particular, the input is a patch of voxels with size $17 \times 17 \times 5$ which upper-bounds the network field of view. We thus limit the network to extracting localized features as compared to commonly used machine learning approaches which aim to incorporate as much global context into the feature extraction process as possible. For example, the U-Net architecture employed in Liu et al. (2019) works with large (496×64) slices of B-scans and comprises three 2×2 pooling operations. On the coarsest scale (bottom of the U), a single convolution with filter size 7×3 thus translates into a field of view of at least 56×24 after unpooling.

5.3 Segmentation via Ordered Assignment

By numerically integrating the ordered assignment flow (2) parametrized by the distance matrix D, an assignment state W is evolved on $\mathscr{W}$ until mean entropy of pixel assignments is low. We specifically use geometric Euler integration steps on $T\mathscr{W}$ with a constant step-length of $h = 0.1$ (see Zeilmann et al. 2020 for details of this process). Geometric averaging with uniform weights leads to local regularization of assignments which smooths regions in which the features do not conclusively point to any label. More global knowledge about the ordering of cell layers is incorporated into E_{ord} which addresses more severe inconsistencies between local features and global ordering. In all experiments, the neighborhood of each voxel $i \in [n]$ is choosen as the voxel patch of size $5 \times 5 \times 3$ centered at i.

5.4 Evaluation

To benchmark our novel segmentation approach, we first extract local features for each voxel from a raw OCT volume. As described above, either region covariance descriptors (Sect. 5.2.1) or class scores predicted by a CNN (Sect. 5.2.3) are computed for segmenting the retina layers with ordered assignment flow which we in the following abbreviate as *OAF (A)* and *OAF (B)* respectively. To facilitate the performance examination between the proposed approach and the reference methods introduced in (Sect. 5.1.2) we evaluate the obtained results through direct comparison of different metrics from (Sect. 5.1.3) and by providing side-by-side visualizations of segmented OCT-volumes in each subsection separately. Specifically, we calculate the DICE similarity coefficient (Dice 1945) and the mean absolute error for segmented cell layers within the pixel size of 3.87 µm compared to human grader by segmenting 8 OCT volumes consisting of 61 B-scans. Throughout the performed experiments, we fixed the grid connectivity $\mathscr{N}_i$ for each voxel $i \in I$ to $3 \times 5 \times 5$.

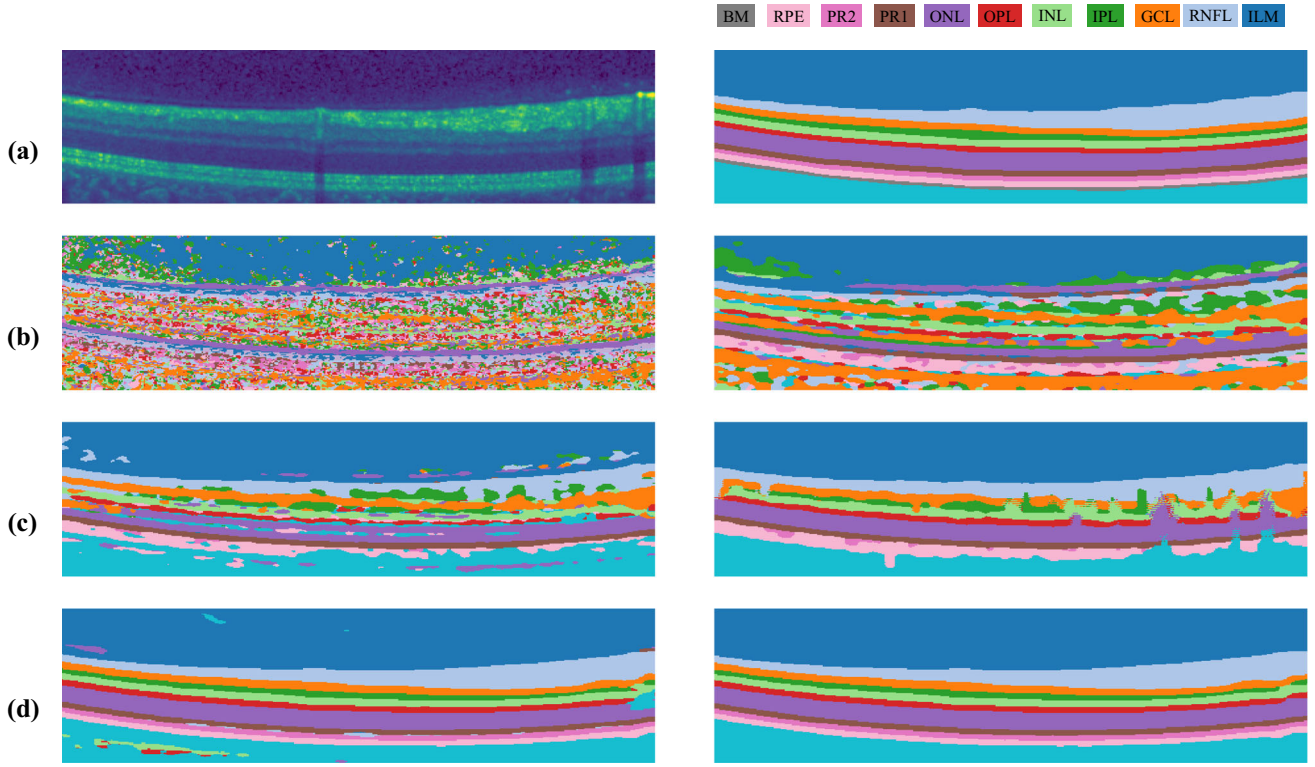


Fig. 7 *From top to bottom: Row* **a** One B-scan from a OCT-volume showing the shadow effect, with ground truth plot on the right. *Row* **b** Local nearest neighbor assignments based on prototypes by minimizing (81) computed with Stein divergence, with the result of the segmentation returned by the basic assignment flow (Sect. 2) on the right. *Row* **c** Proposed *layer-ordered* volume segmentation based on covariance descriptors. From left to right: ordered volume segmentation for different $\gamma = 0.5, \gamma = 0.1$ [cf. Eq. (66)]. *Row* **d** Local rounding result extracted from Res-Net on the left and the result of the ordered assignment flow on the right

5.4.1 Covariance Descriptor vs. CNN

In order to compare *OAF (A)* and *OAF (B)*, we first specifically evaluate the segmentation performance based on local features given by the covariance descriptor (Sect. 5.2.1) as well as features extracted by a CNN (5.2.3). For OAF (A), a dictionary of $k = 400$ prototypical cluster centers on the positive definite cone (Sect. 33) has been determined offline for each retina layer using the iterative clustering with (82). These are compared to descriptors extracted from the unseen volume by computing pairwise Stein divergence (Sect. 3.2.3). The minimum value corresponding to the lowest divergence for each pair of voxel $i \in [n]$ and cell layer $j \in [c]$ is noted as entry d_{ij} of the distance matrix D_{cov}, i.e. for every voxel i the divergence to its closest representative of layer j is given by

$$(D_{\mathrm{cov}})_{ij} := \min_{k \in [400]} D_S(S_i, \tilde{S}_j^k). \tag{87}$$

For OAF (B), class scores $C \in \mathbb{R}^{n \times c}$ predicted by the neuronal network (Sect. 5.2.3) are transformed into a distance matrix $D_{\mathrm{cnn}} = -C$ simply by switching their sign followed by adjusting the parameter ρ to adjust data scale in the likelihood matrix (22).

A naive way to segment the volume in accordance with the observed data is by choosing $\arg\min_{j \in [c]} D_{ij}$ for each voxel i. However, due to the challenging signal-to-noise ratio in real-world OCT data, classes will not usually be well-separated in the feature space at hand. The resulting uncertainty pertaining to the assignment of classes using exclusively local features is encoded into each distance matrix.

The experimental results discussed next illustrate the relative influence of the covariance descriptors (80) and regularization properties of the ordered assignment flow, respectively. To overcome the high computational complexity when extracting features given by (80) and the subsequent assembly of distance matrix (87) during the experiments carried out for *OAF(A)* and *OAF(B)* we segmented OCT volumes consisting of 41 remaining B-scans after cropping 10 B-scans from each volume boundary. Additionally we reduced the size of each B-scan by 148 voxels from each side along the N_A axis to avoid artifacts caused by high varying shape and strong thinning of the retinal layers near volume bounds. Figure 7 illustrates real-world labeling performance based on extracting a dictionary of 400 prototypes per layer by minimizing (81) and employing Algorithm 2 for mean retrieval. The second row in Fig. 7 illustrates a

Table 1 Dice indices (± standard deviation) per cell layer for each of the compared segmentation approaches

	OAF (A)	OAF (B)	Rathke et al. (2014)	IOWA
ILM	0.8837 ± 0.2564	0.9739 ± 0.0189	**0.9972** ± 0.0006	0.9837 ± 0.0043
RNFL	0.6963 ± 0.1998	**0.8842** ± 0.0313	0.8841 ± 0.0125	0.8323 ± 0.0236
GCL	0.6657 ± 0.1909	0.8373 ± 0.0263	**0.8735** ± 0.0152	0.7757 ± 0.0334
IPL	0.5853 ± 0.1773	0.8151 ± 0.0367		0.7860 ± 0.0189
INL	0.6671 ± 0.1773	0.8414 ± 0.0035	0.7501 ± 0.0292	**0.8434** ± 0.0269
OPL	0.7018 ± 0.2013	**0.8442** ± 0.0437	0.7651 ± 0.0124	0.8024 ± 0.0311
ONL	0.8575 ± 0.2523	0.9254 ± 0.0486	**0.9312** ± 0.0068	0.8893 ± 0.0182
PR1	0.8199 ± 0.2407	0.8717 ± 0.0441	0.7945 ± 0.0271	
PR2	0.6787 ± 0.1976	**0.8330** ± 0.0516		
RPE	0.6313 ± 0.1821	**0.8213** ± 0.0835		
CS	0.8606 ± 0.2469	0.9445 ± 0.0488	**0.9858** ± 0.0073	0.9667 ± 0.0167

Lowest mean in bold. The reference methods (Rathke et al. 2014) and IOWA distinguish between a smaller number of cell layers as indicated. Evaluation was performed on a test set consisting of eight OCT volumes (see Appendix C)

Table 2 Mean absolute errors (± standard deviation) per cell layer interface for each of the compared segmentation approaches in pixels (1 pixel = 3.87 μm)

	OAF (A)	OAF (B)	Rathke et al. (2014)	IOWA
ILM-RNFL	1.3590 ± 0.4114	**0.8856** ± 0.3513	1.3080 ± 0.6039	2.7799 ± 0.9485
RNFL-GCL	2.5426 ± 0.7819	**1.4767** ± 0.5589	2.9180 ± 1.0303	2.0561 ± 0.4978
GCL-IPL	3.0183 ± 1.0682	**1.6082** ± 1.5291	–	3.1970 ± 1.1408
IPL-INL	2.6160 ± 1.1294	**1.5004** ± 0.8652	5.1853 ± 1.3642	2.7583 ± 1.3776
INL-OPL	**1.6080** ± 0.5120	1.6220 ± 1.0786	4.8489 ± 1.5898	3.0330 ± 1.2837
OPL-ONL	**1.6342** ± 0.7174	1.8853 ± 1.3951	4.1490 ± 1.2310	4.4292 ± 1.5052
ONL-PR1	**0.6995** ± 0.2467	0.7500 ± 0.3216	–	–
PR1-PR2	**0.6320** ± 0.2442	0.8458 ± 0.4914	5.7281 ± 1.5411	–
PR2-RPE	1.7244 ± 0.6038	**1.2850** ± 1.3660	–	–
RPE-CS	**2.1354** ± 1.0836	2.8613 ± 2.5612	5.2757 ± 1.6452	7.3738 ± 3.2031

Lowest mean in bold. Evaluation was performed on a test set consisting of eight OCT volumes (see Appendix C)

typical result of volume segmentation by nearest neighbor assignment *without* ordering constraint. As expected, the high texture similarity between the choroid and GCL layer yields wrong predictions resulting in violation of physiological cell layer ordering through the whole volume. However, using pairwise correlations captured by covariance matrices leads to accurate detection of the internal limiting membrane (ILM) with its characteristic highly reflective boundary. Similarly, the light rejecting fiber layers RNFL, PR1 and RPE can also be detected by this approach. For the particularly challenging inner layers such as GCL, INL and ONL that are mainly comprised of weakly reflective neuronal cell bodies, regularization by imposing (65) is required. In the third row of Fig. 7, we plot the *ordered* volume segmentation for two different values of the parameter γ defined in (66), which controls the ordering regularization by means of the novel generalized likelihood matrix (68). Direct comparison with the ground truth shows how ordered labelings evolve on the assignment manifold while simultaneously giving accurate data-driven detection of RNFL, OPL, INL and the ONL layer. For the remaining critical inner layers, the local prototypes extracted by (81) fail to represent the retina layers properly and lead to artifacts due to the presence of vertical shadow regions caused by existing blood vessels, which contribute to a loss of the interference signal during the OCT scanning process, as shown in Fig. 7.

After segmentation of the test data set, the mean and standard deviation were calculated for better assessment of the retina layer detection accuracy of the proposed segmentation method, according to the performance measures (75) and (74). The evaluation results for each retina tissue as depicted in Fig. 3, are detailed in Tables 1 and 2. The first row of Fig. 8 clearly shows the superior detection accuracy of utilizing the Ordered Assignment Flow for the first outer retina layers (RNFL, GCL, IPL, INL) and the (PR2-RPE) region in connection with local features extracted by a CNN (Sect. 5.2.3). Nonetheless, the covariance descriptor achieves comparable results for characterization of the outer plexiform layer (OPL) and exhibits increased retina detection regarding the photoreceptor region (PR1,PR2) and outer nuclear

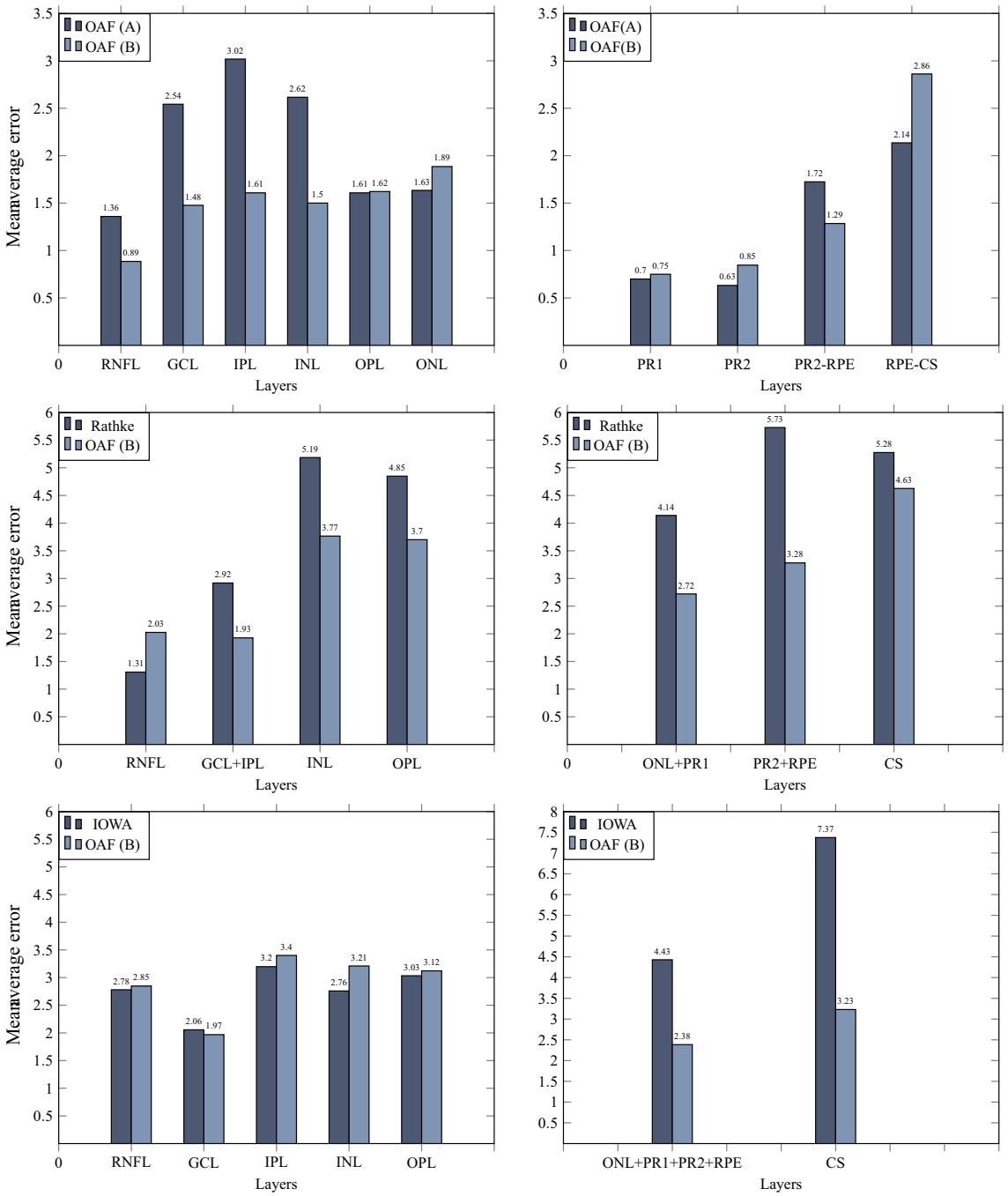


Fig. 8 Performance measures per layer in terms of the mean average error based on the segmentation of 10 healthy OCT volumes. *Top row* Error bars for retina layers separated by the external limiting membrane (ELM) corresponding to OAF (A) and OAF (B). *Middle row* Comparison of the mean errors of OAF (B) and the probabilistic method (Rathke et al. 2014). *Bottom row* Comparison of mean average errors of OAF (B) and the the IOWA reference algorithm

region (ONL). Table 1 includes the evaluation based on Dice similarity which is less sensitive to outliers and serves as an appropriate metric for calculating the performance measures across large 3D volumes. To obtain a consistent and clear comparability between the involved features on which we rely to tackle the specific problem of retina layer segmentation, the corresponding results are visualized in Fig. 9. The graphic illustrates higher Dice similarity and relatively small standard deviation when incorporating features (Sect. 5.2.3) as input to our model, which characterizes their superior informative content. According to the left plot, the covariance descriptor performs well for detecting the prototypical

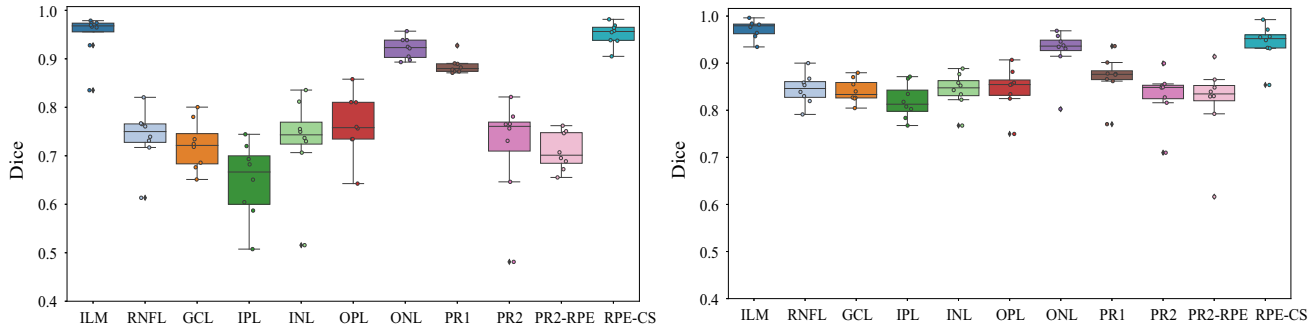


Fig. 9 Box plots of DICE similarity coefficients between computed segmentation results and manually labeled ground truth. *Left* OAF (A). *Right* OAF (B). The OAF based on CNN features yields improved segmentations for all retina layers

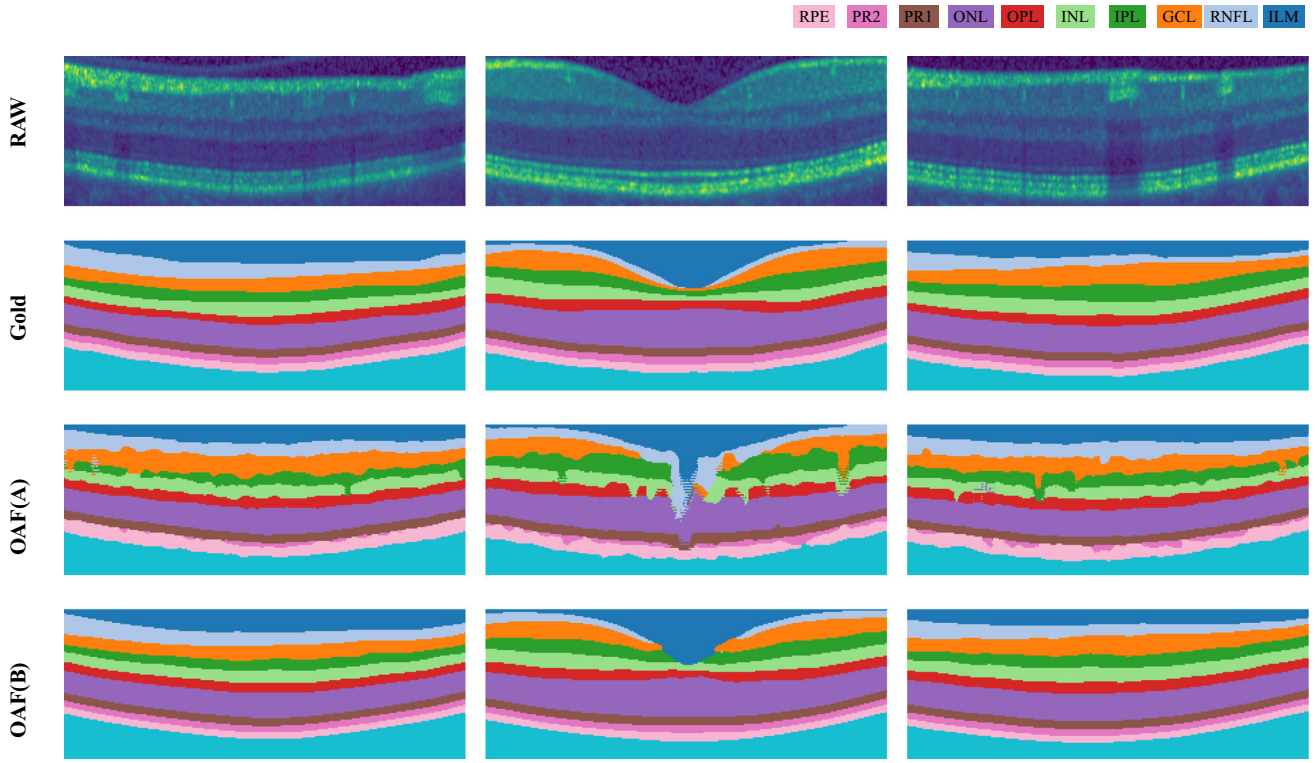


Fig. 10 *From top to bottom* Three sample B-Scans extracted for different locations from a healthy OCT volume with 61 scans, with the fovea centered OCT scan visualized in the middle column. The associated augmented labeling. OAF (A) segmentation using a dictionary of covariance descriptors determined by (82). OAF (B) segmentation using features determined the CNN network. In contrast to to results achieved by OAF (A), the above visualization indicates more accurate detection of retina boundaries using OAF (B), in particular near the fovea region (middle column)

textures of the internal limiting membrane (ILM), the (ONL) and (PR1) layers as well as the RPE boundary to the choroid section. Especially this highlights the ability of using gradient based features for accurate detection of retina tissues indicating sharp contrast between the neighboring layers, as is the case for ONL and PR1.

In general, the more robust retina detection features extracted by a CNN can be attributed to the underlying manifold geometry of symmetric positive definite matrices where the data partition is performed linearly by hyperplanes. This further indicates the nonlinear structure of the acquired volumetric OCT data. Figure 10 presents typical labelings of a B-scan for different locations in the segmented healthy OCT volume obtained with the proposed approach. Direct comparison with the ground truth, as depicted in row (b), demonstrates higher accuracy and smoother boundary transitions by using CNN features instead of covariance descriptors. In particular, for the challenging segmentation

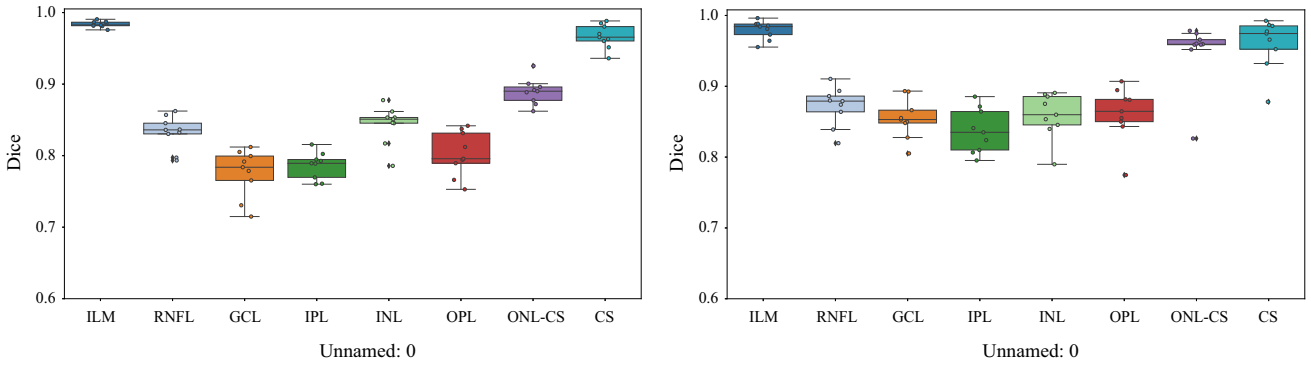


Fig. 11 Box plots of DICE similarity coefficients between computed segmentation results and manually labeled ground truth. *Left* IOWA reference algorithm (Garvin et al. n.d). *Right* OAF based on CNN features. See Table 1 for mean and standard deviations. Exploiting OAF (B) for retina tissue classification results in improved overall layer detection performance, especially for the PR1-RPE region

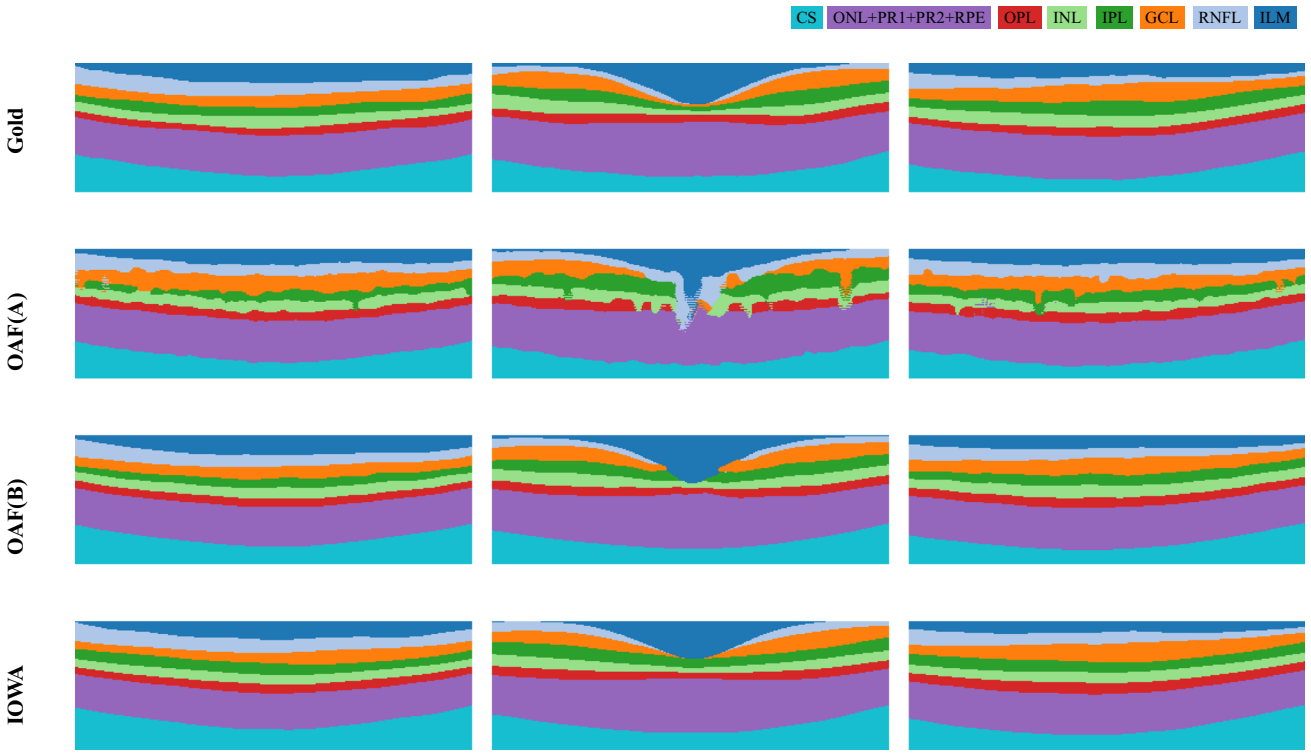


Fig. 12 Illustration of retina layer segmentation results listed in Tables 1 and 2. *From top to bottom* Ground truth labeling. Labeled retina tissues using the proposed approach based on covariance descriptors and CNN features, respectively. The resulting segmentation obtained using the IOWA reference algorithm

of the ganglion cell layer (GCL) with a typical thinning near the macular region (middle scan), we report a Dice index of 0.8373 ± 0.0263 as opposed to 0.6657 ± 0.1909. The remaining numerical experiments are focused on the validation of OAF against the retina segmentation methods summarized in Sect. 5.1.2 serving as reference.

5.4.2 IOWA Reference Algorithm

To assess the segmentation performance of our proposed approach, we first compared to the state of the art graph-based retina segmentation method of 10 intra-retinal layers developed by the Retinal Image Analysis Laboratory at the Iowa Institute for Biomedical Imaging (Kang et al. 2006; Abràmoff et al. 2010; Garvin et al. 2009), also referred to as the IOWA Reference Algorithm. We quantify the region agreement with manual segmentation regarded as gold standard. Since both the augmented volumes and the compared reference methods determine boundary locations of retina layers intersections, we first transfer the retina surfaces to a layer mask by rounding to the voxel size and assign to voxels within each A-scan the associated layer label, starting from

the observed boundary to the location of the next detected intersection surface of two neighboring layers.

To access a quantitative direct comparison with the IOWA reference algorithm, the tested OCT volumes were imported into OCTExplorer 3.8.0 and segmented using the predefined Macular-OCT IOWA software after properly adjusting the resolution parameters. Additionally, we preprocessed each volume by removing 2 B-scans from each side to get rid of boundary artifacts and performed segmentation with the resulting volume size of $498 \times 768 \times 59$ voxels. Quantitative results are summarized in Tables 1 and 2. Figure 11 provides a statistical illustration of the Dice index which reveals the high performance accuracy for methods which is in accordance with the mean average error shown in the last row of Fig. 8. In particular, we observe a notable increase of performance using the OAF for detection of the ganglion cell layer with overall accuracy of $0.8546 \pm 0.0281\,\mu$m, see Fig. 12 for visualized segmentations of 3 B-scans.

5.4.3 Probabilistic Model

Next, we provide a visual and statistical comparison of the proposed approach and the probabilistic state of the art retina segmentation approach (Rathke et al. 2014) underlying Eq. (70). As before, to achieve a direct comparison with the proposed approach, we first adopted the OCT volumes by performing a cropping of 134 voxel from volume boundary along N_A axis to match the shape and parameters for the trained model given in Rathke et al. (2014) which supports the detection of retinal layer boundaries on data sets of dimension $496 \times 500 \times 61$. Subsequently, we removed the boundaries between regions GCL and IPL, ONL and PR1, PR2 and RPE to obtain three characteristic layers which have to be detected. Figure 13 displays the labeling accuracy. Both methods perform well by accurately segmenting flat shaped retina tissues, as shown in the first and last columns. However, closer inspection of the second column reveals a more accurate detection of layer thickness for the (PR2+RPE) and (INL) regions below the concave curved fovea region by using OAF(B). This is mainly due to the connectivity constraints imposed on boundary detection in Rathke et al. (2014). However, the method in Rathke et al. (2014) is more accurate by dealing with rapidly decreasing layer thickness near the fovea region, as observed for GCL and IPL layers in the middle column of Fig. 13 after visual comparison against the manual delineations (first row). In contrast to the Gaussian shape prior used in Rathke et al. (2014), the proposed method does not model connectivity constraints. This allows for the observed oversmoothing artifact, but also makes the OAF approach more amenable for extension to pathological volumes with vanishing retina boundaries. For example, in the case of vitreomacular traction or diabetic macular edema, imposing connectivity constraints aggravates the problem of dealing with irregular retina boundaries.

Figure 14 additionally provides a 3D view on detected retina surfaces for each evaluated reference method used in this publication. The corresponding performance measures (Table 1) underpin the notably higher Dice similarity for (PR2+RPE) and for the (INL) layers. The statistical plots for the mean average error and the Dice similarity index are given in Figs. 8 and 15, clearly showing the overall superiority of OAF (B) with respect to both Dice index and mean average error. In particular, following Table 2, small error rates are observed among all the segmented layers, except for the (ILM) boundary which is detected by all methods with high accuracy. We point out that in general our method is not limited to any number of segmented layers, if ground truth is available.

6 Discussion

We discuss additional aspects pertaining to the data used for training feature extractors as well as the locality of extracted features and limitations of the proposed approach.

6.1 Ground Truth Generation

The training and evaluation of supervised models for feature extraction requires a sizeable amount of high-quality labeled ground truth data. This presents a commonly encountered challenge in 3D OCT segmentation (Dufour et al. 2013; Kang et al. 2006), because the process of manually labeling every voxel of a 3D volume is extremely laborious. The desire to account for inter-observer variability in manual segmentations further compounds this problem. OCT volumes used for testing purposes in the present paper were initially segmented by an automatic procedure based on hand-crafted features. In a subsequent step, each B-scan segmentation was manually corrected by a medical practitioner. The automatic method used for initial segmentation only explicitly regularizes on each individual B-scan, leading to irregularity between consecutive B-scans (see Fig. 16).

Manual correction of initial automatic segmentations leads to a noticeable reduction of irregularity but does not completely remove it. We therefore cannot rule out that a small bias towards the initial automatic segmentation based on hand-crafted features may still be present in the ground truth segmentations that we used to quantify segmentation performance of novel methods as well as baseline methods in this paper. During feature extraction, deep learning models may be capable of discovering the specific hand-crafted features used for initial automated segmentation which may in turn lead to exploitation of any bias towards them. In contrast, because the reference methods are not trained on the

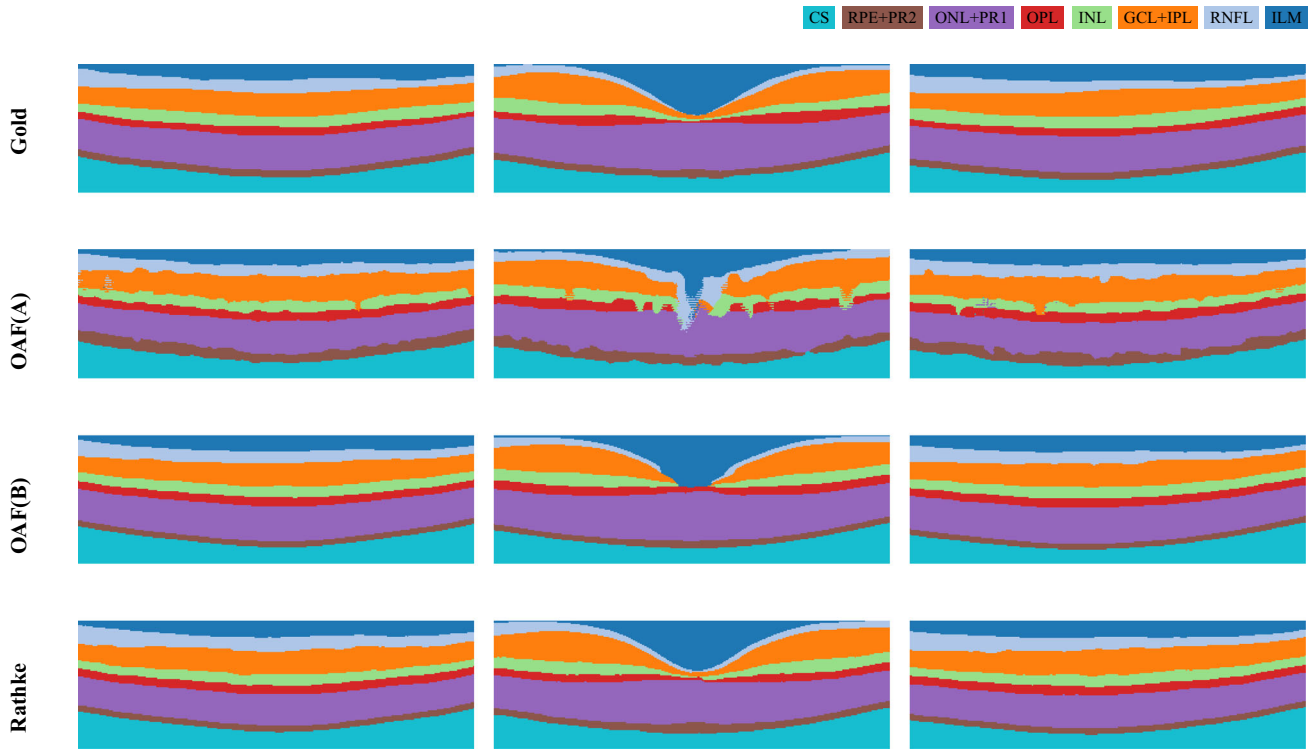


Fig. 13 *From top to bottom* Ground truth for the augmented retina layer corresponding to Table 2. Segmentation results of the OAF based on manifold valued features and on CNN features, respectively. Segmentation results achieved by the probabilistic graphical model approach (Rathke et al. 2014)

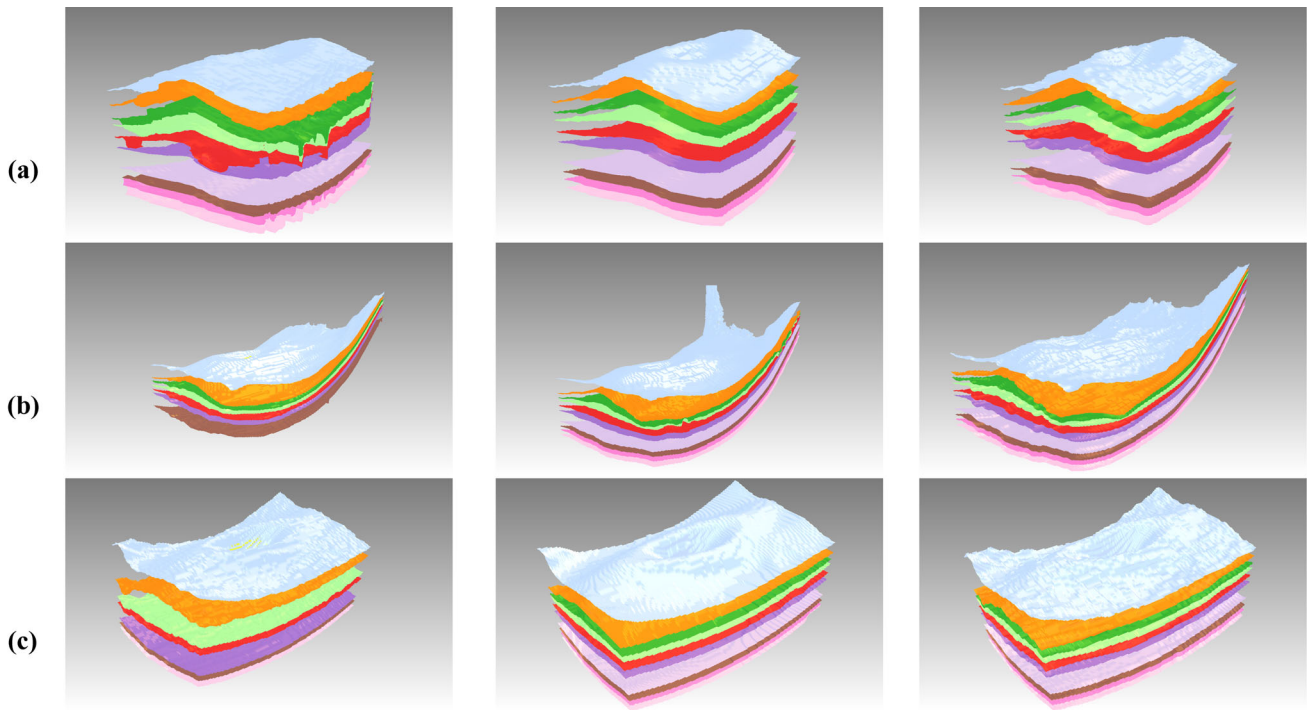


Fig. 14 *Row* **a**: *From left to right:* 3D retinal surfaces determined using OAF (A) (left column) and OAF (B) (middle column). The last column depicts ground truth. Row **b**: *From left to right:* Segmentation of retinal tissues with the IOWA reference algorithm (left column) with the proposed approach (middle column). *Row* **c**: Visual comparison of the probabilistic method (Rathke et al. 2014) (left column) left and the OAF (B) (middle column). Our approach OAF (B) leads to accurate retina layer segmentation with smooth layer boundaries, as observed in the middle column

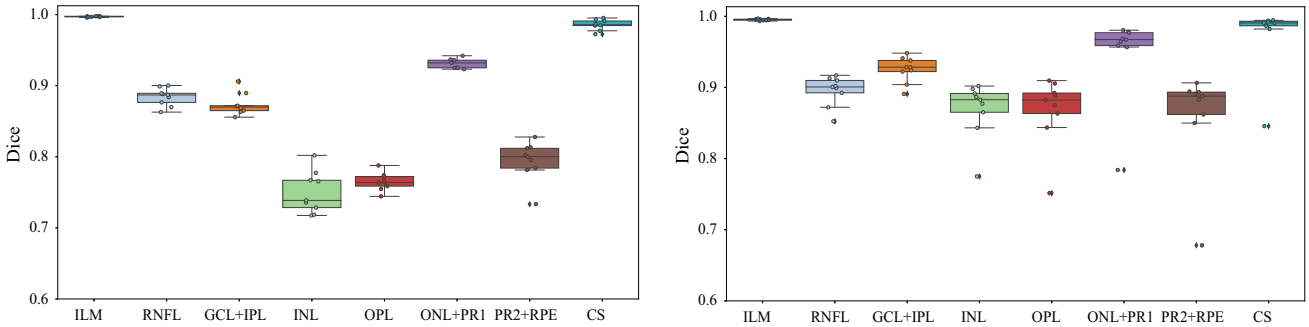


Fig. 15 Box plots of DICE similarity coefficients between computed segmentation results and manually labeled ground truth. *Left* Probabilistic approach (70) proposed in Rathke et al. (2014). *Right* OAF based on CNN features. See Table 1 for mean and standard deviations. Direct comparison shows a notably higher detection performance for segmenting the intraretinal layers using OAF (B)

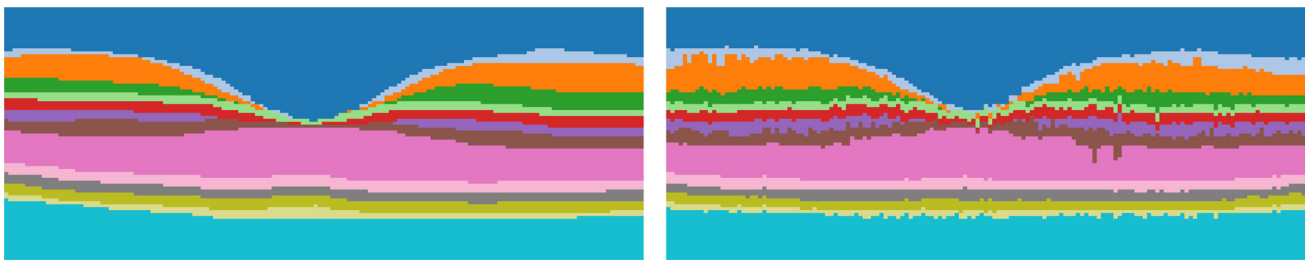


Fig. 16 *Left* Initial automatic segmentation of individual B-scan based on hand-crafted features. *Right* Section of the same automatically segmented volume orthogonal to each B-scan

same data, they can not exploit any such bias, putting them at a possible disadvantage.

Figure 16 also highlights the fact that manual annotations as a gold standard still have nontrivial variance and are partly inconsistent between B-scans. In Rathke et al. (2014), the variance in manual annotations is further analyzed by comparing between two different human observers. They found that for a similar dataset, the discrepancy between both human observers varies between $1.37 \pm 0.51\,\mu\text{m}$ for the most consistent layer boundary and $7.57 \pm 1.06\,\mu\text{m}$ for the least consistent. Comparison to the results in Table 2 (1 pixel $= 3.87\,\mu\text{m}$) illustrates that the proposed model is close to the quality of manual annotation in terms of mean average error. It is to be noted, that similar or even higher scores have been reported for deep learning methods such as Liu et al. (2019) which work on individual B-scans. In view of the inconsistency between manual B-scan segmentations displayed in Fig. 16, it is to be questioned to what extent further improvement of these scores truely reflects improved detection of retina layers if manual annotation is the most precise method available for reference. Part of the contribution of the present work is notably the introduction of a 3D segmentation framework (Definition 2) which serves to regularize by leveraging domain knowledge based on *arbitrary* features. In particular, any deep network can be used as a drop-in replacement for the feature extraction methods discussed in Sect. 5.2.

6.2 Feature Locality

The ordered assignment flow segmentation approach can work with data from any metric space and is hence completely agnostic to the choice of preliminary feature extraction method. In this paper, we chose to limit the field of view of deep networks such that features with local discriminative information are extracted. This makes empirical results directly comparable between features based on covariance descriptors and features extracted by these networks. In addition, we conjecture that local features may generalize better to unseen pathologies. Specifically, if a pathological change in retinal appearance pertains to the global shape of cell layers, local features are largely uneffected. In this way, we expect segmentation performance to be relatively consistent on real-world data. Conversely, widening the field of view in feature extraction should be accompanied by a well-considered training procedure in order to achieve similar generalization behavior, by employing e.g. extensive data augmentation. While raw OCT volume data has become relatively plentiful in clinical settings, large volume datasets with high-quality gold-standard segmentation are not widely available at the time of writing. Therefore, by representing a given OCT scan *locally* as opposed to incorporating global context at every stage, it is our next hypothesis that superior generalization can be achieved in the face of limited data availability. Similarly, although based on local features, the method proposed by Rathke et al. (2014) combines local

knowledge in accordance with a *global* shape prior. This makes clear why some layer scores achieved by this method are very competitive, but it also limits the methods ability to generalize to unseen data if large deviation from the expected global shape seen in training is present.

6.3 Limitations, Future Work

While the OAF typically achieves strong improvement over trivial rounding or baseline regularization, it does not come with a guarantee that physiological layer order will be attained. This is because we use the smooth function (66) instead of the indicator function $\delta_{\mathbb{R}_+^c}$ to define E_{ord} in (65). The parameter γ consequently presents a tradeoff between adherence to physiological layer order and difficulty of numerical integration in the smooth assignment framework (Sect. 2.3). In Fig. 7 [row (c)], this tradeoff becomes apparent when segmenting based on relatively weak covariance descriptor features. Choosing γ smaller leads to improved adherence to the physiological layer order in computed segmentations. However, this also makes numerical integration of the flow (69) more difficult such that the choice of constant step-length $h = 0.1$ may lead to artifacts [row (c), right image]. In such cases, choosing adaptive step-length for integration or using a higher-order numerical integration scheme should still yield stable algorithms at the cost of longer runtime.

We also note that at the fovea, uniformly weighted $5 \times 5 \times 3$ averaging neighborhoods may lead to oversmoothing (see Fig. 13c middle image) which manifests in excessive thinning of e.g. GCL. To combat such artifacts, the choice of averaging weights (23) could be made adaptive to each local neighborhood. However, for most regions of the volume the constant choice of averaging weights made in our experiments does not lead to oversmoothing. Thus, weight adaptivity is to be targeted primarily around the fovea which has a distinctive shape. With regard to computational efficiency, another possible future direction is to encode the notion of layer ordering put forward in Definition 1 within the context of a linear dynamical system for data labeling (Zeilmann et al. 2020).

On the application side, modeling considerations similar to the ones underlying the flow (69) most likely also apply in other areas involving ordering constraints such as seismic horizon tracking for landscape analysis. We thus expect that much of the present work is also relevant outside of optical coherence tomography.

7 Conclusion

In this paper we presented a novel, fully automated and purely data driven approach for retina segmentation in OCT-volumes. Compared to methods (Kang et al. 2006) (Dufour et al. 2013) and (Rathke et al. 2014) that have proven to be particularly effective on tissue classification with a priory known retina shape orientation, our ansatz merely relies on local features and yields ordered labelings which are directly enforced through the underlying geometry of statistical manifold (16). To address the task of leveraging 3D-texture information, we proposed two different feature selection processes by means of region covariance descriptors (80) and the output obtained by training a CNN network as described in Sect. 5.2.3, which are both based on the interaction between local feature responses.

As opposed to other machine learning methods developed for segmenting human retina from volumetric OCT data, the proposed method only takes the pairwise distance between voxels and prototypes (1b) as input. As a direct consequence our approach can be applied in connection with broader range of features living in any metric space and additionally provides the incorporation of outputs from trained neuronal convolution networks interpreted as image features, where a particular instance of such type was demonstrated in Sect. 5.2.3. Even in view of the moderate result achieved after segmentation using OAF (A) in connection with covariance descriptors, we observe the importance of our automatic algorithm by its high level of regularization. Compared to the approach presented in Chiu et al. (2015) which employs a higher number of input features but still requires postprocessing steps to yield order preserving labeling, our approach provides a way to perform this tasks simultaneously.

Using locally adapted features for handling volumetric OCT data sets from patients with observable pathological retina changes is in particular valuable to suppress wrong layer boundaries predictions caused by prior assumptions on retinal layer thicknesses typically made by graphical model approaches as in Dufour et al. (2013) and Song et al. (2013). Our method overcomes this limitation by mainly avoiding any bias towards using priors to global retina shape and instead only relies on the natural biological layer ordering, which is accomplished by restricting the assignment manifold to probabilities that satisfy the ordering constraint presented in Sect. 4. The experimental results reported in Sect. 5, and the direct comparison to the state of the art segmentation techniques (Garvin et al. n.d) and (Rathke et al. 2014) by using common validation metrics, underpin a notable performance and robustness of the geometric segmentation approach introduced in Sect. 2, that we extended to order-preserving labeling in Sect. 4. Furthermore, the results indicate that the ordered assignment flow successfully tackles problems in the field of retinal tissue classification on 3D-OCT data which are typically corrupted by speckle noise, with achieved performance comparable to manual gr-aders which makes it to a method of choice for medical image applications and extensions therein. We point out that our approach consequently differs from common deep learning methods which explicitly aim to incorporate global context

into the feature extraction process. In particular, throughout the experiments we observed higher regularization resulting in smoother transitions of layer boundaries along the B-scan acquisition axis similar to the effect in Rathke et al. (2014) where the used smooth global Gaussian prior leads to limitations for pathological applications.

To reduce the reliance of manually segmented ground truth for extracting dictionaries of prototypes, our method can easily be extended to unsupervised scenarios in the context of Zisler et al. (2020). To deal with highly variable layer boundaries another possible extension of our method is to predict weights for geometric averaging (23) in an optimal control theoretic way, to cope with the linearized dynamics of the assignment flow (Zeilmann et al. 2020) as in detail elaborated in Hühnerbein et al. (2021). Consequently, by building on the feasible concept of spatially regularized assignments (Schnörr 2020), the ordered flow (2) possesses the potential to be extended towards the detection of pathological retina changes and vascular vessel structure. We expect that the joint interaction of retina tissues and blood vessels during the segmentation with the assignment flow will lead to more effective layer detection, which is the objective of our current research.

Acknowledgements We thank Dr. Stefan Schmidt and Julian Weichsel for sharing with us their expertise on OCT sensors, data acquisition and processing. In addition, we thank Fred Hamprecht and Alberto Bailoni for their guidance in training deep networks for feature extraction from 3D data. This work is supported by Deutsche Forschungsgemeinschaft (DFG) under Germany's Excellence Strategy EXC-2181/1 - 390900948 (the Heidelberg STRUCTURES Excellence Cluster).

Funding Open Access funding enabled and organized by Projekt DEAL.

A Proof of Proposition 1

Proof "⇐" Suppose there exists a measure $M \in \mathbb{R}^{c\times c}$ with marginals w_i, w_j and $\langle Q - \mathbb{I}, M\rangle = 0$. Then

$$w_j - w_i = By \Leftrightarrow Q(M - M^\top)\mathbb{1} = y. \tag{88}$$

It suffices to show that no entry of y is negative. Define the shorthand $\zeta = (M - M^\top)\mathbb{1}$. Further, let $M_{\cdot,k}$ denote the k-th column of M and let $M_{k,\cdot}$ denote the k-th row of M. For $l \in [c]$ the components of ζ are given by

$$\zeta_l = (M - M^\top)\mathbb{1}|_l = \langle M_{l,\cdot} - M_{\cdot,l}, \mathbb{1}\rangle = \sum_{k=l}^{c} M_{l,k} - \sum_{k=1}^{l} M_{k,l}. \tag{89}$$

By (88), the entries of y read

$$y_r = \sum_{q=1}^{r} \zeta_q. \tag{90}$$

We can now inductively show that $y_r \geq 0$ for all $r \in [c]$. The cases $r = 1$ and $r = c$ are immediate:

$$y_1 = \zeta_1 = \sum_{k=1}^{c} M_{1,k} - M_{1,1} = \sum_{k=2}^{c} M_{1,k} \geq 0 \tag{91}$$

$$\begin{aligned} y_c &= \langle \zeta, \mathbb{1}\rangle = \langle M - M^\top, \mathbb{1}\mathbb{1}^\top\rangle \\ &= \sum_{i,j\in[c]} M_{i,j} - \sum_{i,j\in[c]} M_{i,j}^\top = 0\,. \end{aligned} \tag{92}$$

For $r \in \{2, \ldots, c-1\}$ we make the hypothesis that

$$y_r = \sum_{q=1}^{r} \zeta_q = \sum_{k=r+1}^{c} \big(M_{1,k} + \cdots + M_{r,k}\big) \geq 0 \tag{93}$$

which is consistent with the result for $r = 1$ in (91). It follows

$$y_{r+1} = \sum_{q=1}^{r+1} \zeta_q \tag{94}$$

$$= \zeta_{r+1} + \sum_{k=r+1}^{c} \big(M_{1,k} + \cdots + M_{r,k}\big) \tag{95}$$

$$= \sum_{k=r+1}^{c} M_{r+1,k} - \sum_{k=1}^{r+1} M_{k,r+1} + \sum_{k=r+1}^{c} \big(M_{1,k} + \cdots + M_{r,k}\big) \tag{96}$$

$$= \sum_{k=r+2}^{c} M_{r+1,k} + \sum_{k=r+2}^{c} \big(M_{1,k} + \cdots + M_{r,k}\big) \tag{97}$$

$$= \sum_{k=r+2}^{c} \big(M_{1,k} + \cdots + M_{r,k} + M_{r+1,k}\big) \tag{98}$$

where we used (93) in (95). This completes the inductive step and thus shows $y \geq 0$.

"⇒" Let (w_i, w_j) be ordered. Following Definition (1), it holds

$$B^{-1}(w_j - w_i) = Q(w_i - w_j) \in \mathbb{R}^c_+. \tag{99}$$

We show the existence of a transport plan $M \geq 0$ satisfying

$$M\mathbb{1} = w_i, \quad M^\top \mathbb{1} = w_j \tag{100}$$

as well as the ordering constraint $\langle Q - \mathbb{I}, M \rangle = 0$ by direct construction. For $c = 2$,

$$M = \begin{pmatrix} (w_j)_1 & (w_i)_1 - (w_j)_1 \\ 0 & 1 - (w_i)_1 \end{pmatrix} \tag{101}$$

satisfies these requirements. Now, let $c > 2$ and define the mapping

$$C_1^{c-1} : \Delta_c \to \Delta_{c-1} \tag{102}$$

$$w \mapsto \tilde{w} = (w_2, \ldots, w_c) + \frac{w_1}{c-1}\mathbb{1}_{c-1}. \tag{103}$$

If $(w_i, w_j) \in \Delta_c^2$ is ordered, then the two assignments

$$(\tilde{w}_i, \tilde{w}_j) := (C_1^{c-1}(w_i), C_1^{c-1}(w_j)) \in \Delta_{c-1}^2 \tag{104}$$

are ordered as well because

$$Q(\tilde{w}_i - \tilde{w}_j) = Q(\bar{w}_i - \bar{w}_j) + \frac{(w_i)_1 - (w_j)_1}{c-1} Q\mathbb{1} \geq 0 \tag{105}$$

where $\bar{w}_i$ denotes the vector $((w_i)_2, \ldots, (w_i)_c)$. Suppose a transport plan $\tilde{M} \in \mathbb{R}^{(c-1)\times(c-1)}$ exists such that

$$\tilde{M}\mathbb{1}_{c-1} = \tilde{w}_i \quad \tilde{M}^\top \mathbb{1}_{c-1} = \tilde{w}_j, \quad \tilde{M} \geq 0. \tag{106}$$

To complete the inductive step, we consider the matrix

$$M^0 := \begin{pmatrix} (w_j)_1 & s^\top \\ 0 & \tilde{M} - \frac{(w_i)_1}{c-1} I \end{pmatrix}, \quad s = \frac{(w_i)_1 - (w_j)_1}{c-1}\mathbb{1}_{c-1} \tag{107}$$

which satisfies (100) as well as $\langle Q - \mathbb{I}, M^0 \rangle = 0$. By Lemma 2, M^0 can be modified to yield a transport plan with the desired properties. □

B Details of employed CNN architecture

As described in Sect. 5.2.3, we employed a CNN architecture for feature extraction which comprises four residually connected blocks. Figure 17 shows a detailed account of how network components are connected. The network produces a sequence of hidden states with channel dimensions

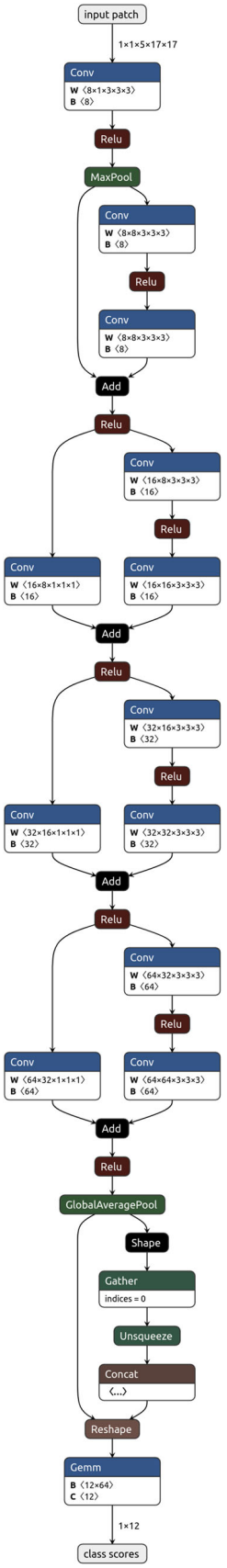


Fig. 17 Convolutional neural network architecture employed for feature extraction

8, 16, 32, 64. Each block contains 3D convolution with filter size $3 \times 3 \times 3$ and rectified linear unit (ReLU) is used as activation function. We trained the network until training loss stopped decreasing after around 4.45M iterations of the stochastic gradient descent optimizer in pytorch with step-length 0.001, momentum 0.9 and batch size 512. Image patches were drawn in random order from the volumes in the training set. During training, we also used dropout with probability 0.3 prior to the single linear layer which decodes class scores.

C Details of Used OCT Data

See Tables 3 and 4.

Table 3 Metadata of OCT volume scans used for training

# B-Scans	# A-Scans	Height (px)	B-Scan distance (μm)	A-Scan distance (μm)	H-Scale (μm/px)
19	1536	496	232.68	5.50	3.87
19	1536	496	249.88	5.51	3.87
19	1536	496	230.57	5.59	3.87
19	1536	496	23.55	5.40	3.87
19	1536	496	241.00	0.58	3.87
19	1536	496	249.07	5.79	3.87
19	1536	496	231.81	5.48	3.87
19	1536	496	255.27	5.78	3.87
19	1536	496	249.04	5.70	3.87
19	1536	496	261.67	5.97	3.87
19	1536	496	244.90	5.58	3.87
19	1536	496	233.74	5.44	3.87
19	1536	496	240.63	5.59	3.87
19	1536	496	236.18	5.45	3.87
19	1536	496	233.71	5.36	3.87
19	1536	496	244.55	0.57	3.87
19	1536	496	252.80	5.93	3.87
19	1536	496	239.38	5.61	3.87
19	1536	496	254.07	0.60	3.87
19	1536	496	247.47	5.83	3.87
19	1536	496	238.06	5.52	3.87
19	1536	496	259.48	6.05	3.87
19	1536	496	26.13	5.88	3.87
19	1536	496	243.29	5.60	3.87
19	1536	496	241.77	5.76	3.87
61	768	496	118.57	11.31	3.87
61	768	496	1.17	11.29	3.87
61	768	496	117.17	11.08	3.87
61	768	496	122.79	11.37	3.87
61	768	496	121.09	11.52	3.87
61	768	496	123.31	11.38	3.87
61	768	496	123.50	11.72	3.87
61	768	496	115.40	10.92	3.87
61	768	496	114.32	10.79	3.87
61	768	496	116.34	10.96	3.87

Table 3 continued

# B-Scans	# A-Scans	Height (px)	B-Scan distance (μm)	A-Scan distance (μm)	H-Scale (μm/px)
61	768	496	119.15	11.30	3.87
61	768	496	127.25	11.81	3.87
61	768	496	126.43	12.19	3.87
61	768	496	121.90	11.45	3.87
61	768	496	12.30	11.64	3.87
61	768	496	124.78	11.86	3.87
61	768	496	123.42	11.40	3.87
61	768	496	120.17	11.40	3.87
61	768	496	126.53	12.04	3.87
61	768	496	115.97	10.96	3.87
61	768	496	128.34	12.19	3.87
61	768	496	124.72	11.74	3.87
61	768	496	119.16	11.10	3.87
61	768	496	119.46	11.23	3.87
61	768	496	123.59	11.80	3.87
61	768	496	118.64	11.07	3.87
61	768	496	125.97	1.21	3.87
61	768	496	119.12	11.47	3.87
61	768	496	122.94	11.65	3.87
61	768	496	129.43	12.07	3.87
61	768	496	116.85	11.26	3.87
61	768	496	122.56	11.64	3.87
61	768	496	128.97	12.09	3.87
512	512	496	0.58	0.58	3.87
512	512	496	0.58	0.58	3.87
256	384	496	11.57	1.15	3.87
256	384	496	11.57	1.15	3.87
512	512	496	0.58	0.58	3.87
256	384	496	11.57	1.15	3.87
512	512	496	5.86	5.85	3.87
256	384	496	11.57	1.15	3.87
512	512	496	5.77	0.58	3.87
256	384	496	11.57	1.15	3.87
512	512	496	0.58	0.58	3.87
512	512	496	0.58	0.58	3.87
256	384	496	12.10	12.05	3.87
512	512	496	6.04	6.03	3.87
512	512	496	6.04	6.03	3.87
512	512	496	6.04	6.03	3.87
512	512	496	5.70	5.69	3.87
19	512	496	242.72	11.38	3.87
19	512	496	241.81	11.33	3.87
19	512	496	242.72	11.38	3.87
19	512	496	242.72	11.38	3.87
19	512	496	241.81	11.33	3.87
19	512	496	245.15	1.15	3.87
19	512	496	242.72	11.38	3.87

Table 4 Metadata of OCT volume scans used for testing

# B-Scans	# A-Scans	Height (px)	B-Scan distance (μm)	A-Scan distance (μm)	H-Scale (μm/px)
61	768	496	116.89	11.08	3.87
61	768	496	120.11	11.33	3.87
61	768	496	123.18	11.72	3.87
61	768	496	127.47	11.94	3.87
61	768	496	127.31	1.23	3.87
61	768	496	122.97	11.52	3.87
61	768	496	113.69	1.10	3.87
61	768	496	124.13	11.80	3.87

References

Abràmoff, M. D., Garvin, M. K., & Sonka, M. (2010). Retinal imaging and image analysis. *IEEE Reviews in Biomedical Engineering, 3*, 169–208.

Amari, S. I., & Nagaoka, H. (2000). *Methods of information geometry*. Amer. Math. Soc. and Oxford Univ. Press.

Antony, B., Abramoff, M., Lee, K., Sonkova, P., Gupta, P., Kwon, Y., Niemeijer, M., Hu, Z., & Garvin, M. (2010). Automated 3-D segmentation of intraretinal layers from optic nerve head optical coherence tomography images. *Progress in Biomedical Optics and Imaging - ProcSPIE, 7626*, 249–260.

Arsigny, V., Fillard, P., Pennec, X., & Ayache, N. (2007). Geometric means in a novel vector space structure on symmetric positive definite matrices. *SIAM Journal on Matrix Analysis and Applications, 29*(1), 328–347. https://doi.org/10.1137/050637996

Åström, F., Petra, S., Schmitzer, B., & Schnörr, C. (2017). Image labeling by assignment. *Journal of Mathematical Imaging and Vision, 58*(2), 211–238.

Bauschke, H. H., & Borwein, J. M. (1997). Legendre functions and the method of random Bregman projections. *Journal of Convex Analysis, 4*(1), 27–67.

Bhatia, R. (2007). *Positive definite matrices*. Princeton University Press.

Bhatia, R. (2013). *The Riemannian mean of positive matrices* (pp. 35–51). Springer.

Bini, D. A., & Iannazzo, B. (2013). Computing the Karcher mean of symmetric positive definite matrices. *Linear Algebra and its Applications, 438*, 1700–1710.

Bridson, M. R., & Häflinger, A. (1999). *Metric spaces of non-positive curvature*. Springer.

Censor, Y. A., & Zenios, S. A. (1997). *Parallel optimization: Theory, algorithms, and applications*. Oxford Univ. Press.

Chan, T. F., & Vese, L. A. (2001). Active contours without edges. *IEEE Transactions on Image Processing, 10*(2), 266–277.

Cherian, A., & Sra, S. (2016). Positive definite matrices: Data representation and applications to computer vision. In H. Minh & V. Murino (Eds.), *Algorithmic advances in Riemannian geometry and applications* (pp. 93–114). Springer.

Chiu, S. J., Allingham, M. J., Mettu, P. S., Cousins, S. W., Izatt, J. A., & Farsiu, S. (2015). Kernel regression based segmentation of optical coherence tomography images with diabetic macular edema. *Biomedical Optics Express, 6*(4), 1172–1194.

Congedo, M., Afsari, B., Barachant, A., & Moakher, M. (2015). Approximate joint diagonalization and geometric mean of symmetric positive definite matrices. *PLoS ONE, 10*(4), e0121423.

Crum, W. R., Camara, O., & Hill, D. L. G. (2006). Generalized overlap measures for evaluation and validation in medical image analysis. *IEEE Transactions on Medical Imaging, 25*(11), 1451–1461.

Depeursinge, A., Foncubierta, A. R., Van De Ville, D., & Müller, H. (2014). Three-dimensional solid texture analysis in biomedical imaging: Review and opportunities. *Medical Image Analysis, 18*(1), 176–196.

Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology, 26*(3), 297–302. https://esajournals.onlinelibrary.wiley.com/doi/pdf/10.2307/1932409

Duan, J., Tench, C., Gottlob, I., Proudlock, F., & Bai, L. (2015). New variational image decomposition model for simultaneously denoising and segmenting optical coherence tomography images. *Physics in Medicine and Biology, 60*, 8901–8922.

Dufour, P. A., Ceklic, L., Abdillahi, H., Schroder, S., De Dzanet, S., Wolf-Schnurrbusch, U., & Kowal, J. (2013). Graph-based multi-surface segmentation of OCT data using trained hard and soft constraints. *IEEE Transactions on Medical Imaging, 32*(3), 531–543.

Fang, L., Cunefare, D., Wang, C., Guymer, R., Li, S., & Farsiu, S. (2017). Automatic segmentation of nine retinal layer boundaries in OCT images of non-exudative AMD patients using deep learning and graph search. *Biomed Optics Express, 8*(5), 2732–2744.

Garvin, M.D., Abràmoff, M.K., & Sonka, M. (n.d). The Iowa Reference Algorithms (Retinal Image Analysis Lab, Iowa Institute for Biomedical Imaging, Iowa City, IA). https://www.iibi.uiowa.edu/oct-reference

Garvin, M. K., Abramoff, M. D., Wu, X., Russell, S. R., Burns, T. L., & Sonka, M. (2009). Automated 3-D intraretinal layer segmentation of macular spectral-domain optical coherence tomography images. *IEEE Transactions on Medical Imaging, 9*, 1436–1447.

Haeker, M., Abramoff, M., Wu, X., Kardon, R., & Sonka, M. (2007). Use of varying constraints in optimal 3-D graph search for segmentation of macular optical coherence tomography images. In *MICCAI* (Vol. 10, pp. 244–51).

Hashimoto, M., & Sklansky, J. (1987). Multiple-order derivatives for detecting local image characteristics. *Computer Vision, Graphics, and Image Processing, 39*(1), 28–55.

He, Y., Carass, A., Liu, Y., Jedynak, B., Solomon, S., Saidha, S., Calabresi, P., & Prince, J. (2019). Deep learning based topology guaranteed surface and MME segmentation of multiple sclerosis subjects from retinal OCT. *Biomedical Optics Express, 10*(10), 5042–5058.

Higham, N. (2008). *Functions of matrices: Theory and computation*. SIAM.

Huang, D., Swanson, E., Lin, C., Schuman, J., Stinson, W., Chang, W., Hee, M., Flotte, T., Gregory, K., & Puliafito, C., et al. (1991). Optical coherence tomography. *Science*, *254*(5035), 1178–1181. https://science.sciencemag.org/content/254/5035/1178.full.pdf

Hühnerbein, R., Savarino, F., Petra, S., & Schnörr, C. (2021). Learning adaptive regularization for image labeling using geometric assignment. *Journal of Mathematical Imaging and Vision, 63*, 186–215.

Jaccard, P. (1908). Nouvelles recherches sur la distribution florale. *Bulletin de la Societe Vaudoise des Sciences Naturelles, 44*, 223–70.

Jost, J. (2017). *Riemannian geometry and geometric analysis* (7th ed.). Springer.

Kafieh, R., Rabbani, H., Abramoff, M., & Sonka, M. (2013). Intra-retinal layer segmentation of 3D optical coherence tomography using coarse grained diffusion map. *Medical Image Analysis, 17*, 907–928.

Kang, L., Xiaodong, W., Chen, D. Z., & Sonka, M. (2006). Optimal surface segmentation in volumetric images-a graph-theoretic approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 28*(1), 119–134.

Kappes, J., Andres, B., Hamprecht, F., Schnörr, C., Nowozin, S., Batra, D., Kim, S., Kausler, B., Kröger, T., Lellmann, J., Komodakis, N., Savchynskyy, B., & Rother, C. (2015). A comparative study of modern inference techniques for structured discrete energy minimization problems. *International Journal of Computer Vision, 115*(2), 155–184.

Kjpargeter, F. (n.d). The muscles of the head. http://www.freepik.com. Accessed 9 Sept 2020

Lee, J. M. (2013). *Introduction to smooth manifolds*. Springer.

Lindeberg, T. (2004). Feature detection with automatic scale selection. *International Journal of Computer Vision, 30*, 79–116.

Liu, X., Cao, J., Fu, T., Pan, Z., Hu, W., Zhang, K., & Liu, J. (2019). Semi-supervised automatic segmentation of layer and fluid region in retinal optical coherence tomography images using adversarial learning. *IEEE Access, 7*, 3046–3061.

Moakher, M., & Batchelor, P. G. (2006). Symmetric positive-definite matrices: from geometry to applications and visualization. In *Visualization and processing of tensor fields* (pp. 285–298). Springer.

Novosel, J., Vermeer, K. A., de Jong, J. H., Wang, Z., & van Vliet, L. J. (2017). Joint segmentation of retinal layers and focal lesions in 3-D OCT data of topologically disrupted retinas. *IEEE Transactions on Medical Imaging, 36*(6), 1276–1286.

Pennec, X., Fillard, P., Ayache, N., & Epidaure, P. (2006). A Riemannian framework for tensor computing. *International Journal of Computer Vision, 66*, 41–66.

Quellec, G., Lee, K., Dolejsi, M., Garvin, M. K., Abramoff, M. D., & Sonka, M. (2010). Three-dimensional analysis of retinal layer texture: Identification of fluid-filled regions in SD-OCT of the Macula. *IEEE Transactions on Medical Imaging, 29*(6), 1321–1330.

Rathke, F., Schmidt, S., & Schnörr, C. (2014). Probabilistic intra-retinal layer segmentation in 3-D OCT images using global shape regularization. *Medical Image Analysis, 18*(5), 781–794.

Rathke, F., Desana, M., & Schnörr, C. (2017). Locally adaptive probabilistic models for global segmentation of pathological OCT scans. *MICCAI* (Vol. 1043, pp. 177–184). Springer.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In*MICCAI* (pp. 234–241). Springer.

Roy, A., Conjeti, S., Karri, S., Sheet, D., Katouzian, A., Wachinger, C., & Navab, N. (2017). ReLayNet: Retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks. *Biomedical Optics Express, 8*(8), 3627–3642.

Schnörr, C. (2020). Assignment flows. In M. Holler, A. Weinmann, & P. Grohs (Eds.), *Variational methods for nonlinear geometric data and applications* (pp. 235–260). Springer.

Sirinukunwattana, K., Snead, D. R., & Rajpoot, N. M. (2015). A novel texture descriptor for detection of glandular structures in colon histology images. *Medical imaging: Digital pathology* (Vol. 9420, pp. 186–194). SPIE.

Sitenko, D., Boll, B., & Schnörr, C. (2020). Assignment flow for order-constrained oct segmentation. In *GCPR* (pp. 58–71).

Song, Q., Bai, J., Garvin, M. K., Sonka, M., Buatti, J. M., & Wu, X. (2013). Optimal multiple surface segmentation with shape and context priors. *IEEE Transactions on Medical Imaging, 32*(2), 376–386.

Sra, S. (2016). Positive definite matrices and the S-divergence. *Proceedings of the American Mathematical Society, 144*(7), 2787–2797.

Turaga, P., & Srivastava, A. (2016). *Riemannian computing in computer vision*. Springer.

Tuzel, O., Porikli, F., & Meer, P. (2006). Region Covariance: A Fast Descriptor for Detection and Classification. In: *Proc. ECCV* (Vol. 3952, pp. 589–600).

Yazdanpanah, A., Hamarneh, G., Smith, B. R., & Sarunic, M. V. (2011). Segmentation of intra-retinal layers from optical coherence tomography images using an active contour approach. *IEEE Transactions on Medical Imaging, 30*(2), 484–496.

Zeilmann, A., Savarino, F., Petra, S., & Schnörr, C. (2020). Geometric numerical integration of the assignment flow. *Inverse Problems, 36*(3), 034004 (33, pp).

Zern, A., Zeilmann, A., & Schnörr, C. (2020a). Assignment flows for data labeling on graphs: Convergence and stability. *CoRR*. arXiv:abs/200211571.

Zern, A., Zisler, M., Petra, S., & Schnörr, C. (2020b). Unsupervised assignment flow: Label learning on feature manifolds by spatially regularized geometric assignment. *Journal of Mathematical Imaging and Vision, 62*(6–7), 982–1006.

Zisler, M., Zern, A., Petra, S., & Schnörr, C. (2020). Self-assignment flows for unsupervised data labeling on graphs. *SIAM Journal on Imaging Sciences, 13*(3), 1113–1156.